



N.º de ordem 07/D/2009

TESE DE DOUTORAMENTO

Apresentada à
Universidade da Madeira

Para obtenção do grau de Doutor

Ana Teresa Gouveia Fernandes

ESTUDO DE MICROSATÉLITES EM HUMANOS: APLICAÇÕES NA IDENTIFICAÇÃO DE INDÍVIDUOS, GENÉTICA DE POPULAÇÕES, DATAÇÃO DE MUTAÇÕES E DIAGNÓSTICO DE DOENÇAS.

ESTUDO DE MICROSATÉLITES EM HUMANOS:
APLICAÇÕES NA IDENTIFICAÇÃO DE INDÍVIDUOS, GENÉTICA DE POPULAÇÕES,
DATAÇÃO DE MUTAÇÕES E DIAGNÓSTICO DE DOENÇAS

Tese de doutoramento apresentada à Universidade da Madeira para a obtenção do grau de Doutor, no Ramo de Ciências Biológicas, Especialidade de Citogenética e Biologia Molecular, sob orientação do Professor Doutor António Manuel Dias Brehm.

Ana Teresa Gouveia Fernandes

Funchal 2009

Ao abrigo do Decreto-lei nº 230/2009, Art. 31º, nº 2, foram compilados na presente dissertação os seguintes trabalhos já publicados ou submetidos para publicação, que serão indicados em numeração romana por ordem da sua referência no texto:

- I. **Fernandes, A.T.**, Brehm, A., Alves, C., Gusmão, L. & Amorim, A. (2002) Genetic profile of the Madeira Archipelago population using the new PowerPlex 16 system kit. *Forensic Science International*, 125, 281-283.
- II. **Fernandes, A.T.**, Brehm, A., Gusmão, L. & Amorim, A. (2001) Y-chromosome STR haplotypes in the Madeira archipelago population. *Forensic Science International*, 122, 178-180.
- III. Velosa, R., **Fernandes, A.T.** & Brehm, A. (2002) Genetic profile of the Açores archipelago population using the new PowerPlex 16 system kit. *Forensic Science International*, 129, 68-71.
- IV. **Fernandes, A.T.** & Brehm, A. (2003) Y-chromosome STR haplotypes in the Açores archipelago (Portugal). *Forensic Science International*, 135, 239-242.
- V. **Fernandes, A.T.**, Velosa, R., Jesus, J., Carracedo, A. & Brehm, A. (2003) Genetic differentiation of the Cabo Verde archipelago population analyzed by STR polymorphisms. *Annals of Human Genetics*, 67, 340-347.
- VI. **Fernandes, A.T.** & Brehm, A. (2002) Population data of five STRs in three regions from Portugal. *Forensic Science International*, 129, 72-74.
- VII. Gonçalves, R., Jesus, J., **Fernandes, A.T.** & Brehm, A. (2002) Genetic profile of a multi-ethnic population from Guiné-Bissau (west African coast) using the new Powerplex 16 system kit. *Forensic Science International*, 129, 78-80.
- VIII. **Fernandes, A.T.**, Rosa, A., Gonçalves, R., Jesus, J. & Brehm, A. (2008a) The Y-chromosome short tandem repeats variation within haplogroups E3b: evidence of recurrent mutation in SNP. *American Journal of Human Biology*, 20, 185-190.
- IX. **Fernandes, A. T.**, Gonçalves, R. & Brehm, A (2008b) Y-chromosome haplotype mismatch in different haplogroups: coincidence or evidence of SNP mutation? *Forensic Science International Genetics*, Suppl1, 200-202.

- X. **Fernandes, A.T.**, Fernandes, S., Gonçalves, R., Sá, R., Costa, P., Rosa, A., Ferrás, C., Sousa, M., Brehm, A. & Barros, A. (2006) DAZ gene copies: evidence of Y chromosome evolution. *Molecular Human Reproduction*, 12, 519-523.
- XI. Berenguer, A., Câmara, R., Brehm, A. & **Fernandes, A. T.** (2009) Distribution of polymorphisms IL4 -590 C/T and IL4 – RP2 in the human populations of Madeira, Azores, Portugal, Cape Verde and Guinea-Bissau. *American Journal of Human Biology (submitted)*.

AGRADECIMENTOS

Ao Professor Doutor António Brehm, orientador deste trabalho, pelo estímulo e confiança durante todos estes anos.

À Universidade da Madeira, nomeadamente ao Departamento de Biologia e ao Laboratório de Genética Humana, pela oportunidade e pela possibilidade de concretizar este trabalho.

A todos os co-autores dos artigos que fazem parte integrante deste trabalho, e respectivas instituições, por toda a colaboração.

Aos meus colegas do Laboratório de Genética Humana que me ajudaram a realizar todos os trabalhos com um grande espírito de equipa.

Aos meus amigos por todos os momentos partilhados especialmente ao João e à Maria João pelas longas noites em que trabalhámos juntos.

Aos meus filhos António, Mariana e Matilde pela atenção e tempo que lhes roubei.

Ao Roberto por estar sempre presente e por me apoiar incondicionalmente.

Aos meus pais e aos meus irmãos pelo apoio ao longo da vida e em especial à minha irmã Susana pela partilha de experiências e cooperação na investigação.

LISTA DE ABREVIATURAS

AR - Androgen Receptor
AZF - Azoospermic Factor
DAZ - Deleted in Azoospermia
DMSO - Dimerhysulphoxide
DNA - Deoxyribonucleic Acid
dNTP - Deoxynucleoside Triphosphate
FMR1 - Fragile X Mental Retardation
IAM - Infinite Allele Model
ISFG - International Society of Forensic Genetics
ISFH - International Society of Forensic Haemogenetics
KAM - K Allele Model
LD - Linkage Disequilibrium
MSI - Microsatellite Instability
mt DNA - Mitochondrial DNA
nm - nanómetros
OR - Odds Ratio
pb - Pares de Bases
PCR - Polymerase Chain Reaction
RFLP - Restriction Fragment Length Polymorphism
SMM - Stepwise Mutation Model
SNP - Single Nucleotide Polymorphism
SSM - Slip Strand Mispairing
STR - Short Tandem Repeat
Taq - *Thermus aquaticus*
TPM - Two Phase Model
UCO - Unequal Crossing-Over
VNTR - Variable Number of Tandem Repeat

RESUMO

Os microsatélites, também chamados STRs (*Short Tandem Repeat*), são pequenas sequências de DNA que consistem numa sequência de repetições de um motivo que varia de um a seis pares de bases. Existem em quase todos os cromossomas humanos e podem situar-se nos exões ou nos intrões. Estes últimos são altamente polimórficos e são por isso utilizados na identificação de indivíduos em testes de paternidade e também em estudos de genética de populações. A combinação dos vários genótipos possíveis faz com que cada indivíduo possua um perfil único, que permite a sua identificação. Existem também microsatélites associados a exões ou a regiões promotoras dos genes, normalmente repetições trinucleotídicas CGG/CCG ou CAG/CTG, associados a doenças neurodegenerativas como a síndrome do X-frágil e a doença de Huntington.

Neste trabalho caracterizaram-se geneticamente várias populações humanas dos arquipélagos da Madeira, Açores e Cabo Verde.

A partir do estudo dos microsatélites do cromossoma Y, foram definidas idades de coalescência que permitiram concluir que as cópias do gene *DAZ* situado no cromossoma Y são o resultado de um processo evolutivo estando a sua evolução associada a alguns haplogrupos. Verificou-se também a ocorrência de possíveis mutações nos SNPs que definem os haplogrupos, através da comparação dos microsatélites do cromossoma Y dentro de cada haplogrupo, especialmente no haplogrupo E3b.

Verificou-se existir uma associação entre o número de repetições CAG e GGC do gene Receptor de Androgénios (AR), situado no cromossoma X, e a infertilidade especialmente quando combinados os dois polimorfismos, parecendo haver um efeito protector dos alelos maiores e alguma susceptibilidade para os alelos menores.

Quando se estudou o número de repetições GGC do gene *FMR1* em doentes com suspeita de síndrome de X-frágil observaram-se diferenças significativas quando comparadas com a população em geral e com um grupo de sobredotados. Essa diferença deveu-se principalmente à presença do alelo 29 em quase todos os indivíduos do primeiro grupo o que por si só não constitui um factor de risco mas poderá ser uma indicação da associação deste alelo com outra mutação no mesmo gene que possa ser responsável por este fenótipo.

ABSTRACT

Microsatellites, also known as STRs (Short Tandem Repeat), are small DNA sequences made out of a series of repetitions of a motive varying from one to six base pairs. They exist in almost all human chromosomes and they can exist either in the exons or in the introns. These introns are highly polymorphic and are therefore very frequently used as a means of identification in paternity tests and population studies. A characteristic combination of the various possible genotypes gives each person a unique profile, which allows identification. There are also microsatellites associated to exons or in the gene promoter regions, usually trinucleotide repetitions CGG/CCG or CAG/CTG, associated to neurodegenerative diseases like X-Fragile and Huntington disease.

In this work various human populations of the archipelagos of Madeira, the Azores and Cape-Verde were genetically characterizes.

From the study of the Y-chromosome microsatellites, coalescence ages were defined which allowed us to conclude that DAZ-gene copies located in the Y-chromosome are the result of an evolutionary process, with its development being associated to specific haplogroups. It was verified that possible mutations were taking place amongst the SNPs that define haplogroups. This was verified by comparing the Y-chromosome microsatellites within each haplogroup, especially in haplogroup E3b.

An association exists between the numbers of CAG and GGC repeats in the Androgen Receptor (AR), in the X-chromosome, and infertility, especially when both are combined, and there seems to exist a protective mechanism for the larger alleles and some susceptibility within the smaller alleles.

When the number of GGC-repeats in gene FMR1 was studied in suspected patients of X-fragile Syndrome, significant differences were found when compared a general population group and one with high QI. This difference was mainly due to the presence of allele 29 in almost all individuals from the former group. This does not represent, *per si*, a risk-factor, but it could be a clue for an association between this allele and a mutation in the same gene that would be responsible for this phenotype.

ÍNDICE

| | |
|--|---------------|
| 1. INTRODUÇÃO | 1 |
| 1.1 Os microsatélites | 3 |
| 1.1.1 Mecanismo de aparecimento dos microsatélites | 4 |
| 1.1.2 Mecanismos de reparação nos microsatélites | 7 |
| 1.1.3 Taxa de mutação dos microsatélites | 8 |
| 1.1.4 Os microsatélites autossómicos | 11 |
| 1.1.5 Os microsatélites do cromossoma Y | 12 |
| 1.1.6 Os microsatélites do cromossoma X | 14 |
| 1.1.7 Homoplasia | 15 |
| 1.1.8 Instabilidade de microsatélites (MSI) | 16 |
| 1.1.9 O Desequilíbrio de Ligação (LD - <i>Linkage Disequilibrium</i>) | 17 |
| 1.1.10 Nomenclatura dos microsatélites | 19 |
| 1.2 Aplicações dos microsatélites | 21 |
| 1.2.1 Estudos populacionais | 22 |
| 1.2.2 Testes de paternidade e identificação de indivíduos | 25 |
| 1.2.3 Associação com doenças | 26 |
| 1.2.3.1 Expansão de trinucleótidos | 27 |
| 1.2.3.2 Perda de heterozigotia e instabilidade de microsatélites | 29 |
| 1.2.3.3 Variação do tamanho dos alelos no Receptor dos Androgéneos | 31 |
| 1.3 História das populações | 33 |
| 1.3.1 Arquipélago da Madeira | 33 |
| 1.3.2 Arquipélago dos Açores | 35 |
| 1.3.3 Arquipélago de Cabo Verde | 36 |
| 1.4 Objectivos | 37 |
| 2. MATERIAL E MÉTODOS | 39 |
| 2.1 Selecção das amostras | 41 |
| 2.1.1. Estudos Populacionais | 41 |

| | |
|--|----|
| 2.1.2. Estudo da variabilidade de haplótipos no cromossoma Y | 43 |
| 2.1.3. Estudo da variabilidade do gene <i>DAZ</i> do cromossoma Y | 43 |
| 2.1.4. Detecção de trissomias por PCR | 43 |
| 2.1.5. Repetições CAG/GGC do gene AR, associadas à infertilidade masculina | 44 |
| 2.1.6. Repetições CAG/GGC do gene AR, associado aos níveis de PSA | 44 |
| 2.1.7. Determinação do número de repetições GGC do gene <i>FMR1</i> | 44 |
| 2.1.8. Estudo da variabilidade de dois polimorfismos do gene <i>IL4</i> | 45 |
| 2.2. Extração de DNA | 45 |
| 2.3. Quantificação do DNA | 45 |
| 2.4. Amplificação do DNA por PCR | 46 |
| 2.4.1. Microsatélites autossómicos | 46 |
| 2.4.2. Detecção de trissomias por PCR | 46 |
| 2.4.3. Microsatélites do cromossoma Y | 47 |
| 2.4.4. Estudo da variabilidade do gene <i>DAZ</i> do cromossoma Y | 49 |
| 2.4.5. Repetições CAG/GGC no gene receptor dos Androgéneos (AR) | 49 |
| 2.4.6. Repetições CGG do gene <i>FMR1</i> | 50 |
| 2.4.7. Estudo da variabilidade de dois polimorfismos do gene <i>IL4</i> | 50 |
| 2.5. Visualização do DNA | 51 |
| 2.5.1. Electroforese em gel de acrilamida | 51 |
| 2.5.2. Electroforese capilar | 51 |
| 2.6. Análise estatística | 51 |
| 3. RESULTADOS | 59 |
| 3.1 Estudos populacionais | 61 |
| 3.1.1 População da Madeira | 61 |
| 3.1.2 População dos Açores | 62 |
| 3.1.3 População de Cabo Verde | 63 |
| 3.2 Bases de dados populacionais | 63 |
| 3.2.1 Bases de dados dos microsatélites autossómicos | 63 |
| 3.2.2 Bases de dados dos microsatélites do cromossoma Y | 65 |
| 3.3 Associação a doenças | 66 |

| | |
|---|------------|
| 3.3.1 Detecção de trissomias por PCR | 66 |
| 3.3.2 Repetições GGC e CAG do gene AR em indivíduos com níveis de PSA elevados | 66 |
| 3.3.3 Repetições GGC e CAG do gene AR em indivíduos com infertilidade masculina | 67 |
| 3.3.4 Comparação entre o número de repetições CAG e GGC do gene AR | 69 |
| 3.3.5 Repetições GGC no gene FMR1 para a população da Madeira e Guiné-Bissau | 70 |
| 3.4 análise de mutações | 73 |
| 3.4.1 Cromossoma Y | 74 |
| 3.4.2 Gene DAZ | 74 |
| 3.4.3 Gene IL4 | 75 |
| 4. DISCUSSÃO E CONCLUSÕES | 77 |
| 4.1 Estudos populacionais | 80 |
| 4.2 Bases de dados e testes de paternidade | 81 |
| 4.3 Associação a doenças | 81 |
| 4.4 Análise de mutações | 83 |
| 4.5 Perspectivas futuras | 84 |
| 5. REFERÊNCIAS BIBLIOGRÁFICAS | 85 |
| 6. PUBLICAÇÕES | 111 |

ÍNDICE DE FIGURAS

| | |
|--|----|
| Figura 1: Mecanismo de <i>slippage</i> durante a replicação de DNA | 7 |
| Figura 2: Localização dos vários microsátélites no cromossoma Y | 13 |
| Figura 3: Mapa de localização dos arquipélagos da Madeira, Açores e Cabo Verde | 33 |
| Figura 4: Alelos GGC do gene FMR1 em três subpopulações da Madeira | 73 |

ÍNDICE DE TABELAS

| | |
|--|----|
| Tabela 1: Microsátélites autossómicos para a determinação das trissomias | 47 |
| Tabela 2: Referências dos microsátélites do cromossoma Y utilizados | 48 |
| Tabela 3: <i>Primers</i> para a amplificação das repetições CAG e GGC do gene AR | 49 |
| Tabela 4: Fórmulas para o cálculo do índice de paternidade (trio) | 56 |
| Tabela 5: Fórmulas para o cálculo do índice de paternidade (filho e pressuposto pai) | 57 |
| Tabela 6: Resultados para um caso de paternidade com o pressuposto pai e um filho | 64 |
| Tabela 7: Cálculo de IP e W com bases de dados diferentes | 65 |
| Tabela 8: Repetições GGC e CAG do gene AR, em indivíduos com PSA elevado | 67 |
| Tabela 9: Repetições GGC e CAG do gene AR, associado a infertilidade masculina | 68 |
| Tabela 10: Comparação CAG e GGC para indivíduos inférteis e férteis | 69 |
| Tabela 11: Comparação CAG e GGC em indivíduos com níveis de PSA elevados | 70 |
| Tabela 12: Repetições GGC do gene FMR1 para a Madeira e Guiné-Bissau | 71 |
| Tabela 13: Repetições GGC do gene FMR1 para três grupos da população da Madeira | 72 |

1. INTRODUÇÃO

1.1 OS MICROSATÉLITES

Os microsátélites, também chamados STRs (Short Tandem Repeat; Edwards *et al.*, 1991a), são pequenas sequências de DNA que consistem em repetições de um motivo que varia de um a seis pares de bases (Sutherland e Richards, 1995; Schlötterer, 2000).

Os microsátélites são majoritariamente constituídos por cadeias poli-A e poli-T e apesar de se encontrarem em quase todos os organismos, são mais comuns em organismos com genomas maiores, aparecendo com frequências diferentes consoante o genoma e evidenciando uma excepcional variabilidade nos humanos (Hancock, 1999). Muitos microsátélites originalmente tipados em humanos são já polimórficos em macacos (Coote e Bruford, 1996) havendo descrições de microsátélites com milhões de anos nas tartarugas e peixes (Fitzsimmons *et al.*, 1995; Rico *et al.*, 1996).

A variabilidade faz dos microsátélites marcadores genéticos de eleição para muitas aplicações, incluindo mapeamento genético e estudos evolutivos que permitem determinar relações entre espécies e entre populações, para além de uma indiscutível importância em genética médica, forense e da conservação. Os trinucleótidos mais frequentemente encontrados são o CAG e AAT (Stallings, 1994) e os dinucleótidos mais comuns são os CA e TG ocorrendo duas vezes com mais frequência do que as repetições AT e 3 vezes mais do que as AG e TC (Beckman e Weber, 1992). Os dinucleótidos não são bons marcadores para estudos de populações devido à sua alta instabilidade durante a PCR (*Polymerase Chain Reaction*), fazendo com que a determinação do tamanho dos alelos se torne difícil ou mesmo impossível.

Apesar de haver outros *loci* com maiores taxas de mutação, os microsátélites têm sido utilizados para estabelecer métodos moleculares que permitam definir mecanismos e taxas de mutação. Os SNPs (*Single Nucleotide Polymorphism*) têm a vantagem de ser

fáceis de detectar e de tipar mas a grande desvantagem é que devido à sua baixa taxa de mutação a maior parte ocorre uma vez só, levando a que estes polimorfismos sejam bialélicos (Armour, 2006).

Mesmo nas mutações que ocorrem em *loci* neutrais, a alteração no tamanho das repetições em *tandem* serve para ilustrar mecanismos como a replicação, recombinação ou reparação pois os agentes que causam essas mutações podem promover da mesma maneira outras mutações no genoma. Por outro lado a variação no número dessas repetições pode também produzir alterações no fenótipo (Armour, 2006).

A designação de satélites deve-se ao facto de durante a 1ª experiência em 1961 (quando se falavam dos satélites como o “Sputnik” nas notícias) algumas fracções menores terem formado bandas separadas do grosso do DNA tendo sido por isso chamadas de satélites em relação à fracção maior. Essas fracções representavam os VNTR (*Variable Number of Tandem Repeat*) incluindo os que normalmente se encontram nos centrómeros dos cromossomas das várias espécies (Jobling *et al.*, 2004).

1.1.1 Mecanismo de aparecimento dos microsatélites

Uma mutação é qualquer alteração na sequência de DNA, podendo variar entre uma substituição de uma só base, ou pequenas inserções e deleções de algumas bases. A maior parte das mutações nos genes humanos nas linhas germinais devem-se ao resultado de vários processos endógenos envolvendo mecanismos químicos (metilação e desaminação da 5-metilcitosina nos dinucleótidos CpG), físicos (DNA *slippage*) ou enzimáticos (falhas no mecanismo de reparação pós-replicativas e exonuclease *proofreading*) (Cooper *et al.*, 1993).

A maior parte das diferenças observadas entre o genoma de vários indivíduos não têm qualquer influência fenotípica. Estas diferenças designam-se de mutações neutrais (Jobling *et al.*, 2004). Se as mutações ocorrerem nas linhas germinativas podem passar de geração em geração, se não forem letais e forem compatíveis com a fertilidade. As mutações nas células somáticas podem ter consequências graves muito especialmente

em humanos (como o cancro) mas não têm importância do ponto de vista evolutivo (Jobling *et al.*, 2004).

Desde a descoberta da estrutura do DNA em 1953 (Watson e Crick, 1953), foram apresentadas várias explicações para o aparecimento de mutações devido a erros durante a replicação do DNA sendo as substituições de bases não complementares ou a adição ou deleção de um ou vários nucleótidos durante a síntese do DNA os mecanismos propostos para o aparecimento de alterações na sequência do DNA (Fresco e Alberts, 1960).

Nos microsatélites o mecanismo predominante para o aparecimento de mutações deve-se à falha no emparelhamento de cadeias de DNA durante a replicação podendo igualmente ocorrer durante a recombinação. Durante a replicação do DNA há um desalinhamento das duas cadeias de DNA (DNA *slippage*) (Schlötterer e Tautz, 1992) e se não houver uma reparação e a síntese do DNA prosseguir, o número de repetições altera-se podendo aumentar ou diminuir (Levinson e Gutman, 1987; Strand *et al.*, 1993). À medida que o número de repetições aumenta a taxa de *slippage* tende a aumentar pois a probabilidade de desemparelhamento também aumenta (Tachida e Iizuka, 1992).

As sequências repetitivas ocorrem nos genomas dos eucariotas através de mecanismos comuns nomeadamente mecanismos de *slippage*, e não parecem ter nenhuma função em especial na expressão de genes. É de esperar que se encontrem por todo o genoma principalmente em zonas não sujeitas a selecção tendo, no entanto, uma função indirecta de criar zonas cruciais para a recombinação (Tautz e Renz, 1984), sendo bastante comuns nos centrómeros da maioria dos cromossomas (Lee *et al.*, 1997). Um grande número de microdeleções e microinserções no DNA humano aparecem como uma consequência de mecanismos de *slippage* durante a replicação mediados pela presença de repetições directas ou invertidas (palindrómicas) nas suas imediações (Cooper *et al.*, 1993).

O processo de reparação do DNA evita que estas alterações ocorram mais frequentemente (Schlötterer, 2000). Os mecanismos de reparação durante o emparelhamento corrigem desemparelhamentos e curvas no DNA (*loops*) que provocam inserções ou deleções ocorrendo espontaneamente durante processos metabólicos do DNA como a replicação e a recombinação (Fishel e Wilson, 1997; Kolodner, 1996).

Estudos efectuados em bacteriófagos demonstram que deleções e inserções de unidades de repetição ocorrem frequentemente em regiões repetitivas sendo mais frequente a sua ocorrência em sequências maiores, variando apenas uma unidade de repetição (Levinson e Gutman, 1987).

Rose e Falush (1998) demonstraram a necessidade da existência de um tamanho mínimo para uma sequência repetitiva estabelecer mutações dinâmicas mas quando ocorre uma expansão existe uma maior probabilidade de ocorrência de novas expansões pois aumenta a possibilidade da ocorrência de *slippage* adicionais (Levinson e Gutman, 1987).

Existem vários modelos que podem justificar o aparecimento de diferentes tamanhos nos microsatélites: o modelo IAM - *Infinite Allele Model* (Kimura e Crown, 1964); SMM - *Stepwise Mutation Model* (Kimura e Ohta, 1978); TPM - *Two Phase Model* (Di Rienzo *et al.*, 1994) e KAM - *K Allele Model*, que corresponde ao IAM quando K é infinito e parece ser mais realista do que o IAM.

Segundo o modelo IAM cada mutação produz um novo alelo que é diferente dos já existentes podendo resultar em qualquer número de repetições enquanto no modelo SMM o tamanho dos microsatélites varia com uma taxa fixa independente do tamanho da repetição e com a mesma probabilidade de expansão e contracção (Valdes *et al.*, 1993). Os modelos IAM e SMM representam dois modelos de mutação extremos (Chakraborty e Jin, 1992). Microsatélites com repetições de 3 a 5 pares de bases evoluirão segundo o modelo SMM (Shriver *et al.*, 1993) mas os microsatélites com motivos de repetição menores evoluirão através do modelo MSM - *Multiple-Step Mutation* (Di Rienzo *et al.* 1994), mais próximo do modelo IAM.

Se ocorrer um desemparelhamento sem posterior reparação, o resultado será diferente se este ocorrer na cadeia molde ou na cadeia nascente, dando origem a uma repetição a mais ou a menos consoante a cadeia afectada (figura 1).

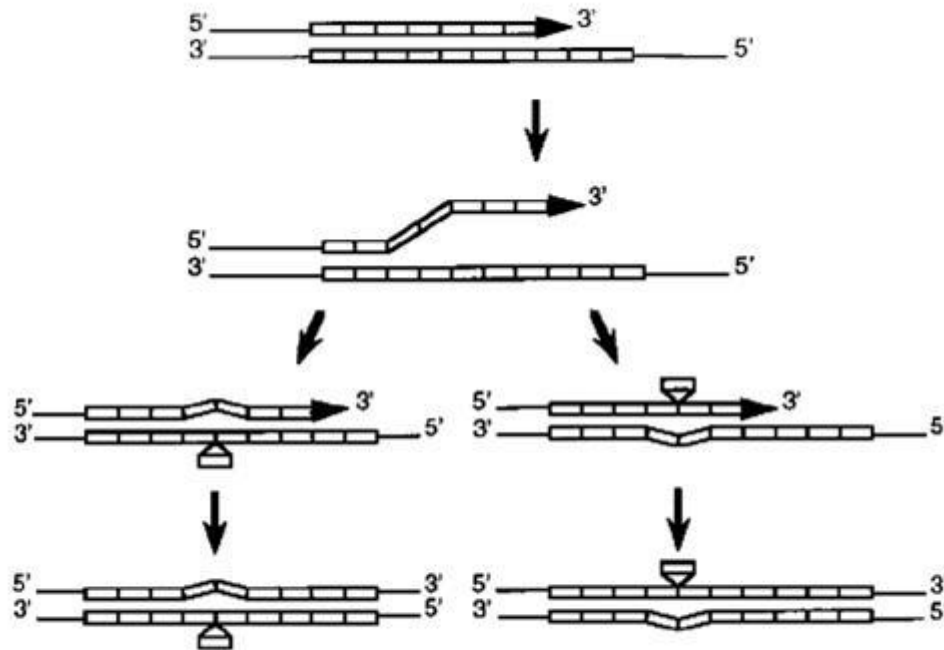


Figura 1: Mecanismo de *slippage* durante a replicação de DNA. Se houver um desemparelhamento sem reparação posterior observa-se a formação de um *loop* que se ocorrer na cadeia molde originará a perda de uma repetição (lado esquerdo) e se ocorrer na cadeia nascente originará o ganho de uma repetição (lado direito) (Sia *et al.*, 1997).

1.1.2 Mecanismos de reparação nos microsatélites

Muitos erros originados pela *Taq* DNA polimerase não se tornam em mutações devido ao mecanismo de correcção das exonucleases que degradam as novas cadeias replicadas quando há erros, fazendo com que a DNA polimerase retroceda e a cadeia seja recopiada (Kunkel, 1992).

Este mecanismo de replicação é limitado na reparação de mutações nos microsatélites visto que estas exonucleases apenas reconhecem os erros que ocorrem a poucas bases de distância do sítio activo da polimerase. Como muitos dos *loops* gerados por SSM (*Slip Strand Mismatching*) estão longe do sítio de replicação acabam por não ser reconhecidos pelas exonucleases, o que justifica a elevada taxa de mutação dos microsatélites quando comparada com a taxa das mutações pontuais. Este mecanismo funciona de diferentes

maneiras no interior dos microsatélites resultando melhor quando estes são pequenos e quando a unidade de repetição é curta (Eisen, 1999).

O mecanismo de reparação de falhas (*mismatch*) permite fazer o reconhecimento e reparação de alterações na sequência de bases. O mesmo processo pode reparar o DNA contendo *loops* como os gerados pelo mecanismo SSM nos microsatélites. Este mecanismo é muito mais importante na estabilidade dos microsatélites do que o das exonucleases. O papel do mecanismo de reparação de *loops* nos microsatélites é bem comprovado no caso do cancro do cólon não poliposo em que a instabilidade de microsatélites se deve a um defeito hereditário neste mecanismo de reparação (Eisen, 1999).

Há dois complexos de proteínas MutS homóloga (MSH) para a reparação do DNA nuclear, o MSH2-MSH6 que reconhece e repara as bases e *loops* que têm entre 1 a 2 bases e o MSH2-MSH3 que reconhece e repara *loops* com 1 a 6 bases. Defeitos noutra proteína MutS homóloga (MSH1) causa o aumento da taxa de mutação no DNA mitocondrial apesar destes mecanismos ainda serem pouco conhecidos. Há ainda outras proteínas (MSH4 e MSH5) que não têm funções reparadoras mas sim de regulação do *crossing-over* durante a meiose (Eisen, 1999).

A reparação depende da natureza dos microsatélites nomeadamente do seu tamanho: *loops* de 1 a 3 bases são reparados facilmente, com 4 bases são mal reparadas e com mais de 4 não são reparadas devido à incapacidade das proteínas MutS reconhecerem *loops* maiores do que 4 bases (Parker e Marinus, 1992).

1.1.3 Taxa de mutação dos microsatélites

A taxa de mutação e a estrutura dos microsatélites é determinada pela combinação da ocorrência do tipo de erros SSM e de como esses erros são reparados. Por exemplo, uma mutação pode ocorrer com uma frequência elevada porque deriva de um erro SSM comum ou porque é mal reparada. Apesar de erros SSM envolvendo duas ou mais repetições não serem frequentes, a maior parte das mutações em microsatélites com repetições de 5 pb resultam em alterações de duas ou mais repetições porque as

alterações de uma só repetição são reparadas enquanto as outras não. Uma vez que o número e o tamanho das repetições influenciam a estabilidade dos microsátélites é importante comparar repetições com o mesmo tamanho quando se estuda o efeito do número de cópias e manter o número de repetições quando se pretende estudar o efeito do tamanho da unidade de repetição (Eisen, 1999).

A taxa de mutação dos microsátélites pode ser afectada por vários factores como a probabilidade de desalinhamento, a probabilidade da continuação da polimerização sem correcções no alinhamento e a eficácia da reparação dos heteroduplexes, sendo a sua diminuição uma das razões para o aparecimento de instabilidade de microsátélites (Kunkel, 1993).

A taxa de mutação aumenta com o número de repetições pois há mais sítios de ocorrência de *slippage* e desemparelhamento (Eisen, 1999) mas este aumento não é regular pois as repetições perfeitas são significativamente mais variáveis do que as repetições imperfeitas, o que pode explicar porque se observa uma correlação positiva entre a taxa de mutação e o número de repetições em alguns estudos e noutros não (Goldstein e Clark, 1995).

A taxa de mutação nos microsátélites é muito menor que a dos minisátélites mas nestes depende do tamanho e do número de repetições e principalmente do comprimento de repetições perfeitas enquanto a instabilidade dos minisátélites não depende nem do tamanho nem do facto das repetições serem perfeitas (Armour, 1999).

As taxas de mutação variam consoante os microsátélites estudados devendo-se a maior parte das mutações ao ganho ou perda de uma só repetição (Hoff-Olsen *et al.*, 1998). A taxa de mutação dos microsátélites é de cerca de 10^{-3} por *locus*, por gâmeta e por geração (Brinkmann *et al.*, 1998; Weber e Wong, 1993) enquanto a taxa de mutação para substituição de nucleótidos é cerca de $2,5 \times 10^{-8}$ mutações por nucleótido (Nachman e Crowell, 2000).

A ocorrência de mutações nos gâmetas masculinos é cerca de seis vezes maior do que nos gâmetas femininos aumentando com a idade paterna uma vez que as espermatogónias se renovam continuamente por mitose antes de se transformarem em espermatozóides enquanto as oogónias se dividem apenas cerca de 22 vezes antes do início da meiose que dará origem aos oócitos. Desta forma a variação nos microsátélites é

causada essencialmente pelos homens (Brinkmann *et al.*, 1998). São mais frequentes as mutações envolvendo ganhos de uma repetição do que perdas e estas ocorrem 4 vezes com mais frequência nas células germinais masculinas do que nas femininas (Weber e Wong, 1993; Amos *et al.*, 1996) devido ao número de replicações que durante a oogénese é constante enquanto na espermogénese varia com a idade e ocorre cerca de 14 vezes mais do que na oogénese (Jobling *et al.*, 2004).

As mutações no interior dos próprios microsatélites são frequentes, alterando o seu comprimento devido a inserções e deleções de pequenas unidades de repetição (Weber e Wong, 1993), aumentando com o aumento do tamanho dos alelos (Xu *et al.*, 2000). A média da taxa de mutação para os microsatélites tetranucleotídicos é 4 vezes maior do que para os dinucleotídicos (Weber e Wong, 1993; Goldstein e Pollock, 1997) embora haja estudos que indiquem uma maior taxa para estes últimos (Chakraborty *et al.*, 1997; Zhivotovsky *et al.*, 2000). A verdade é que estes estudos baseiam-se em métodos diferentes sendo o primeiro resultado da observação directa das mutações e o segundo derivado de estudos populacionais (Estoup e Cornuet, 1999).

A ocorrência de transições é duas vezes mais frequente que a ocorrência de transversões dado que o mecanismo de reparação do DNA é mais sensível a estas últimas (Jobling *et al.*, 2004).

As taxas de mutação para transições ou transversões em dinucleótidos CpG são maiores do que as taxas de mutação noutras zonas e as substituições de um só nucleótido ocorrem 10 vezes mais do que mutações maiores (Nachman e Crowell, 2000) pois a hipermetilação dos dinucleótidos CpG despoleta nas linhas germinais transições para TG e CA com frequências 5 a 7 vezes maiores do que as taxas de mutações pontuais (Krawczak *et al.*, 1998).

As substituições são dez vezes mais frequentes que as inserções e deleções (*indels*) (Jobling *et al.*, 2004) e nos microsatélites há um favorecimento de expansões observado em vários estudos: 21 expansões contra 9 contracções (Amos *et al.*, 1996) e 26 expansões contra 7 contracções (Primmer *et al.*, 1996), apesar de haver outros estudos que mostram que as mutações ocorrem simetricamente (Di Rienzo *et al.*, 1994; Goldstein *et al.*, 1995).

Existem dois modelos que podem explicar as altas taxas de mutação dos microsatélites: o primeiro envolve apenas um desemparelhamento na cadeia dupla

durante a replicação (Levinson e Gutman, 1987) enquanto o segundo envolve recombinação entre moléculas de DNA quer por *crossing-over* quer por conversão génica. O *crossing-over* desigual leva a que uma cadeia fique com uma deleção e outra com uma inserção e pode ocorrer entre cromátides do mesmo cromossoma ou entre cromossomas. A conversão génica, que envolve transferência de informação unidireccional por recombinação, provavelmente como resposta a uma danificação do DNA, pode gerar diversidade nos microsatélites (Jeffreys *et al.*, 1994)

A principal evidência do papel do mecanismo de *slippage* na ocorrência de mutações é o aparecimento de variações por perda ou ganho de apenas uma unidade de repetição (Hancock, 1999).

1.1.4 Os microsatélites autossómicos

Os microsatélites autossómicos têm sido descritos desde 1991 e estão situados em vários *loci* de autossomas diferentes. Um dos primeiros a ser descrito foi o *locus* CD4 situado no braço curto do cromossoma 12 e é um polimorfismo pentanucleotídico (CTTTT). O primeiro trabalho onde este foi descrito apresentou frequências para duas populações distintas: uma de caucasianos e outra de negros norte americanos (Edwards *et al.*, 1991b). Outro polimorfismo tetranucleotídico foi apresentado para o locus TH01 situado no cromossoma 11, região p15.5 (Polymeropoulos *et al.*, 1991) sendo posteriormente reavaliada a designação de alguns dos seus alelos (Puers *et al.*, 1993). Edwards e colaboradores apresentaram em 1991 um sistema multiplex que permitiu a amplificação simultânea de 3 microsatélites tri e tetranucleotídicos nos cromossomas 4, 11 e X, estudando a variação destes polimorfismos em 4 populações dos Estados Unidos (Edwards *et al.*, 1991a). Em 1992 foram publicadas pela *International Society of Forensic Haemogenetics* (ISFH) as recomendações para o uso de polimorfismos de DNA em ciências forenses aplicada à tecnologia de PCR (Bar, 1992). Foi só a partir de 1996 que surgiram os primeiros *kits* comerciais de amplificação simultânea de vários microsatélites, existindo actualmente *kits* para 15 microsatélites bem como para a determinação do sexo através do gene da amelogénina (AmpF Φ STR®IdentifilerPCR Amplification Kit, Applied

Biosystem; PowerPlex® 16 System, Promega). Através do uso destes microsatélites é possível proceder à identificação de indivíduos e de relações de parentescos bem como efectuar estudos populacionais com a finalidade de estudar principalmente o grau de miscigenação das populações bem como a frequência dos alelos dentro dos vários microsatélites que aparecem muitas vezes associados a populações com origens étnicas diferentes (Dupalomp e Bertorelle, 2001).

Devido à elevada variabilidade dos microsatélites é possível estudar outras mutações recorrendo ao estudo de desequilíbrio de ligação (LD - *Linkage Disequilibrium*) existentes entre os haplótipos formados por vários microsatélites e essas mutações. No entanto esta associação depende fortemente do número de duplos heterozigóticos visto que nesse caso as combinações podem ser múltiplas e não é possível utilizar esses indivíduos para o estudo (Stephens *et al.*, 1999).

1.1.5 Os microsatélites do cromossoma Y

O cromossoma Y, por ser haploíde, não recombina no processo meiótico, à excepção da região homóloga do cromossoma X, sendo herdado directamente de pais para filhos (Jobling e Tyler-Smith, 1995). O acumular de mutações representa um registo da evolução das populações pois os polimorfismos permitem inferir sobre os movimentos migratórios masculinos. Desde 1995 que a análise deste cromossoma é utilizada para traçar e comparar linhagens paternas (Jobling e Tyler-Smith, 1995; Hammer *et al.*, 1997).

Em 2001, Stumpf e Goldstein sugerem a utilização de dois métodos para determinar genealogias utilizando microsatélites do cromossoma Y, fazendo simulações de coalescências em que mostram que o processo de mutação pode alterar as conclusões que se inferem (Stumpf e Goldstein, 2001).

Em 1997, Kayser e colaboradores apresentam o estudo de 13 microsatélites (DYS19, DYS288, DYS385, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, YCAI, YCAII, YCAIII, DXYS156Y) (Kayser *et al.*, 1997) localizados em várias zonas do cromossoma Y (figura 2). Os primeiros estudos populacionais com microsatélites do cromossoma Y aparecem nessa altura com o estudo de uma população Italiana (Caglià *et al.*, 1997). É

apresentada uma nomenclatura consistente com o número de repetições para cada um dos *loci* e uma estimativa da taxa de mutação para o DYS19 de $3,2 \times 10^{-3}$ que se prevê ser semelhante a outros microsátélites com estrutura idêntica (Kayser *et al.*, 1997). Quase todas as mutações são de apenas uma repetição, sendo mais frequentes os ganhos do que as perdas (Kayser *et al.*, 2000). Os alelos longos são mais mutáveis e a taxa de mutação parece aumentar com o aumento da idade paterna (Gusmão *et al.*, 2005).

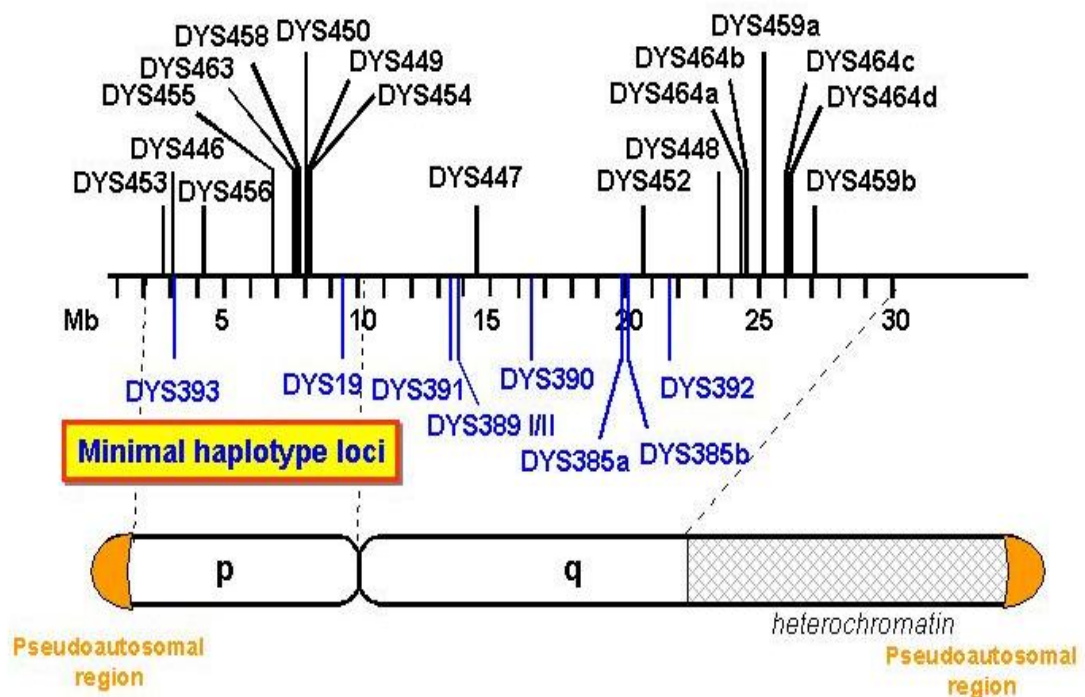


Figura 2: Localização dos vários microsátélites no cromossoma Y

(<http://www.cstl.nist.gov/div831/strbase/ystrpos1.htm>)

A definição de haplótipo é dada como a combinação dos alelos de dois ou mais genes localizados no mesmo cromossoma (Jobling *et al.*, 2004).

Os microsátélites do cromossoma Y quando utilizados em conjunto formam haplótipos que podem ser utilizados para identificação de linhagens paternas a exclusão de uma paternidade ou a identificação de um indivíduo. Contudo, estes haplótipos não servem de confirmação de paternidades ou identificação individual uma vez que o mesmo haplótipo é, em princípio, partilhado por todos os membros masculinos da família descendentes de

um mesmo indivíduo. Um exemplo é o caso descrito por Junge e colaboradores (2006) em que, apesar de terem sido utilizados 22 STRs autossômicos, apenas foram encontradas duas exclusões para o pressuposto pai, tendo-se por isso recorrido aos haplótipos do cromossoma Y. Como estes eram diferentes, concluiu-se então que se tratava de uma exclusão de paternidade (Junge *et al.*, 2006).

Numa mistura, se a proporção entre a amostra do macho em relação à fêmea for baixa, o perfil do macho apenas poderá ser detectado através da análise de marcadores do cromossoma Y como os microsatélites (Gusmão *et al.*, 2006; Prinz *et al.*, 1997). Os haplótipos formados pelos vários microsatélites do cromossoma Y são um bom auxiliar em genética forense pois permitem avaliar a significância da observação de um determinado haplótipo numa amostra (Pereira *et al.*, 2002).

Existe actualmente uma base de dados mundial (<http://www.yhrd.org>) representativa de 477 populações, com 54,863 haplótipos formados pelo menos por 8 microsatélites do cromossoma Y, designado por haplótipo mínimo (Willuweit e Roewer, 2007).

Utilizando um grande número de microsatélites do cromossoma Y podem obter-se resultados mais precisos de idade de coalescência e mais compatíveis com eventos migratórios recentes (King *et al.*, 2007).

1.1.6 Os microsatélites do cromossoma X

As mutações ligadas ao cromossoma X estão num estado haplóide 1/3 do tempo, logo a variabilidade da sua *fitness* é maior do que as mutações autossômicas, fazendo com que a selecção se torne mais eficaz e o aumento da frequência mais rápida para as mutações benéficas ligadas ao X (Avery, 1984). Isto significa que os *loci* ligados ao X devem apresentar um número mais elevado de alelos raros do que os autossomas, visto que o cromossoma X tem níveis de variabilidade mais baixos quer para os microsatélites quer para os SNPs, parecendo ser o efeito da selecção mais acentuado no cromossoma X do que nos autossomas (Payseur *et al.*, 2002).

Os maiores desvios na distribuição das frequências encontram-se predominantemente em regiões de baixa recombinação sendo este efeito mais pronunciado no cromossoma X

(Payseur *et al.*, 2002). A distribuição espacial dos vários marcadores pode fornecer informação útil na identificação de regiões do genoma sujeitas a selecção. Uma discrepância nas frequências entre o mtDNA e os *loci* autossómicos pode ser causada por um efeito gargalo (*bottleneck*), o qual é menos perceptível no cromossoma sexual na medida em que a sua relação com os autossomas é de 3/4 enquanto com o mtDNA é de apenas 1/4 (Payseur *et al.*, 2002).

Em 1990 foram descritos polimorfismos para 5 microsatélites constituídos por repetições de 2 pares de bases, no cromossoma X (Luty *et al.*, 1990). Actualmente várias populações de origens étnicas diferentes têm sido estudadas utilizando um conjunto de 10 STR do cromossoma X numa só reacção (Gomes *et al.*, 2007; Gusmão *et al.*, 2009).

Diferenças entre os sexos em factores demográficos como a taxa de migração, podem deixar diferentes marcas no cromossoma X e nos autossomas. Esta discrepância pode indicar uma taxa de fixação alta ou um tempo de transição baixo para mutações benéficas no cromossoma X (Avery, 1984).

1.1.7 Homoplasia

A homoplasia resulta da convergência de alelos diferentes em alelos com o mesmo tamanho embora uma parte dessas mutações nos microsatélites possam produzir alelos idênticos em tamanho mas diferentes na sequência. Esta possibilidade faz com seja necessário ter algum cuidado ao utilizar microsatélites em estudos filogenéticos ou outros, em que os indivíduos estudados estejam separados por um grande número de gerações mas com um ancestral comum (Garza e Freimer, 1996). Estas convergências tanto podem ser causadas por mutações numa só região repetitiva ou por mutações em duas zonas repetitivas homólogas (Di Rienzo *et al.*, 1994).

A homoplasia é uma consequência do modelo SMM (*Stepwise Mutation Model*) e resulta em alelos idênticos em linhagens diferentes (Cooper *et al.*, 1996).

A homoplasia nos microsatélites não representa um problema para a maior parte dos estudos de genética de populações pois a grande variabilidade dos microsatélites compensa a ocorrência de homoplasia durante a sua evolução. As situações onde a

homoplasia pode ser mais problemática envolvem altas taxas de mutação e populações grandes associados com elevadas limitações ao tamanho dos alelos (Estoup *et al.*, 2002). Segundo o modelo SMM, a probabilidade de ter ao acaso dois alelos com o mesmo tamanho é igual à homozigotia (Kimura e Ohta, 1978).

Alguma da variabilidade dos microsatélites fica a dever-se às regiões flanqueadoras da região repetitiva. É comum observar alelos com o mesmo tamanho mas com diferentes estruturas internas que estão separadas por zonas flanqueadoras que apresentam muita variação, fenómeno que indica que o processo mutacional que origina os polimorfismos numa região envolve não só simples alterações no número de repetições mas também alterações nas zonas flanqueadoras. Há no entanto casos de homoplasias em microsatélites que não envolvem diferenças nessas zonas (Grimaldi e Crouau-Roy, 1997).

Quando se consideram haplótipos baseados em múltiplos *loci* a extensão da homoplasia é reduzida mas existe sempre a possibilidade de um determinado haplótipo ter várias origens (Cooper *et al.*, 1996).

1.1.8 Instabilidade de microsatélites (MSI)

Existem dois modelos propostos para o mecanismo de instabilidade de microsatélites. Ambos justificam, embora de maneiras diferentes, o mecanismo de ocorrência de mutações. Um dos modelos, UCO (*Unequal Crossing-Over*), propõe que a instabilidade é o resultado da recombinação entre cromossomas homólogos, causada por uma elevada taxa de *crossing-over* entre as repetições de microsatélites (Smith, 1976). O modelo alternativo, SSM (*Slip Strand Mismatching*), propõe que a instabilidade é causada pela elevada taxa de desemparelhamento durante a replicação do DNA (Eisen, 1999). Este processo foi proposto por Fresco e Alberts em 1960 para explicar a ocorrência de mutações *frameshift* em todo o DNA (Fresco e Alberts, 1960). O mecanismo começa durante a replicação com o *slipping* da DNA polimerase fazendo com que a cadeia molde e as novas cadeias fiquem momentaneamente desalinhadas. Para que a replicação continue as cadeias devem realinhar-se. As mutações acontecem quando o realinhamento é imperfeito. Este modelo parece ser o mais provável para explicar a

instabilidade nos microsátélites uma vez que a presença de repetições aumenta a tendência de desalinhamento após a *slippage* visto que as repetições podem facilmente criar *loops* fora da dupla cadeia (Streisinger *et al.*, 1966).

Os desalinhamentos ocorrem mais frequentemente em regiões repetitivas do que no DNA normal visto que os *loops* gerados pelo desalinhamento são mais estáveis nos microsátélites do que em regiões não repetitivas pois o emparelhamento de bases não é significativamente alterado quando uma ou mais cópias de uma repetição estão em *loop* (Eisen, 1999).

O processo SSM pode ser subdividido em três passos: 1) *slipping* da DNA polimerase durante a replicação; 2) desalinhamento da cadeia molde e das novas cadeias e 3) continuação da replicação apesar do desalinhamento das cadeias.

O *slippage* ocorre frequentemente nos microsátélites visto que estes são mais propícios à formação de estruturas de DNA menos usuais existindo uma maior taxa de SSM em regiões contendo microsátélites do que noutras regiões do genoma que pode ser causada pelo aumento da ocorrência quer de *slippage* quer de desalinhamento das cadeias durante a replicação (Eisen, 1999). A ocorrência de SSM varia também consoante os microsátélites estando correlacionado com o número de cópias da repetição onde fragmentos de repetições com mais de 51 pares de bases sofrem significativamente mais deleções do que fragmentos com menos do que 33 pares de bases e consequentemente para fragmentos maiores a ocorrência de adições é mais frequente do que para fragmentos pequenos em que ocorrem igualmente deleções e adições (Wierdl *et al.*, 1997).

1.1.9 O Desequilíbrio de Ligação (LD - *Linkage Disequilibrium*)

Por vezes verificam-se combinações haplotípicas com maior prevalência do que a esperada tendo em conta a frequência dos alelos que a compõem. Neste caso diz-se que existe Desequilíbrio de Ligação (LD - *Linkage Disequilibrium*) o qual consiste numa associação não casual de dois ou mais alelos de diferentes *loci* no mesmo cromossoma.

O LD é a associação não aleatória de alelos de diferentes *loci* fazendo com que a definição de um dos alelos seja redundante quando se trata de mapeamento de genes. Factores como a selecção, mutações locais e taxas de recombinação podem influenciar o LD em regiões particulares do genoma. Alterações no tamanho das populações e a mistura entre populações diferentes podem igualmente influenciar o padrão de LD no genoma (Goldstein, 2001).

O LD é, ao nível populacional, a dependência de vários alelos situados em diferentes posições do genoma. Quando ocorre recombinação entre dois *loci* tende a haver uma redução da dependência entre os alelos nesses *loci* e isso reduz o LD (Li e Stephens, 2003).

Um efeito associado a um determinado alelo de um microsatélite pode ser, de facto, o resultado de um LD com outra variante nomeadamente SNPs por vezes situados a alguma distância do gene em questão (Armour, 2006).

A diversidade no DNA em humanos resulta de mutações nas linhas germinativas e ocorre devido a recombinação meiótica (Jeffreys e Neumann, 2002). Essas mutações não ocorrem ao acaso mas apenas em alguns sítios que constituem pontos críticos (*hotspots*). Existem por isso muitas regiões do genoma humano que constituem haplótipos em bloco de 5 a 100 Kb (Daly *et al.*, 2001) sendo que os limites desses haplótipos são normalmente pontos críticos (Jeffreys e Neumann, 2002). Não há contudo evidências de que todos os limites correspondam a pontos críticos devendo ser a diversidade dos haplótipos simultaneamente influenciada pela história demográfica das populações (Anderson e Slatkin, 2004).

Um menor LD pode significar a existência de recombinação ou tratar-se de uma população mais antiga, logo sujeita a mais fenómenos evolutivos de recombinação. Numa população com tamanho constante a existência de pontos críticos de recombinação pode resultar numa baixa diversidade dentro dos haplótipos em bloco mas esta pode também dever-se a um crescimento rápido da população. Esta é, aliás, a principal razão para o aparecimento de haplótipos em bloco e não a existência de pontos críticos de recombinação. A existência de blocos diferentes contendo SNPs em perfeito ou quase perfeito LD significa que a recombinação entre os cromossomas contendo os diferentes haplótipos é baixa (Anderson e Slatkin, 2004), contrariamente ao descrito por Slatkin

(Slatkin, 1994) que mostra que há menos LD numa população de rápido crescimento do que quando o crescimento é constante. A explicação para esta discrepância baseia-se no facto de no primeiro caso as diferenças entre haplótipos terem sido determinadas antes do crescimento da população enquanto no segundo caso as simulações foram feitas assumindo que todas as variações surgiram por mutação depois do crescimento rápido começar (Anderson e Slatkin, 2004).

Alguns estudos mostram que os padrões de LD estão altamente estruturados em grandes blocos e que apenas existem alguns pontos críticos de quebra. Dentro dos blocos não existem evidências de recombinação e a diversidade haplotípica é baixa apesar da taxa de recombinação no genoma ser elevada, sendo mais frequente nos telómeros e menos nos centrómeros (Stumpf e Goldstein, 2003).

A existência de blocos discretos contendo um LD consistente poderá reduzir a necessidade de inferências sofisticadas na relação gene-população no mapeamento genético. As comparações a fazer deverão ser no sentido de procurar recombinações discretas raras ou populações mais antigas onde o LD é menor, devido ao aumento da probabilidade de mutação e recombinação mesmo dentro dos blocos, sendo relevante descrever as estruturas haplotípicas para várias populações (Goldstein, 2001).

Para análises de LD são necessários trinta vezes mais SNPs do que microsatélites visto estes serem muito mais polimórficos do que os primeiros (Havill e Dyer, 2005), sendo por isso recomendado o uso de microsatélites para a determinação de LD (Ohashi e Tokunaga, 2003).

1.1.10 Nomenclatura dos microsatélites

Diz-se que há um polimorfismo quando existem pelo menos dois alelos presentes na população com uma frequência mínima de 1% (Cummings, 2000).

Desde 1992 foram publicadas recomendações acerca da utilização de polimorfismos de DNA em ciências forenses. Devido ao facto de algumas zonas do DNA serem codificantes e outras não, os segmentos de DNA polimórficos (ex D1S80, D17S5, HLA-DQ) foram designadas por *locus*. O termo alelo deverá ser utilizado para descrever formas

alternativas de um *locus*. Sempre que possível a designação D (designação do *Human Gene Mapping*) deverá ser utilizada para descrever o *locus* e a numérica para definir os alelos (Bar *et al.*, 1992).

A nomenclatura dos alelos dos microsatélites deve ser feita segundo o número de repetições completas. Se existir uma repetição parcial a designação é feita pelo número completo de repetições separado por um ponto seguido do número de bases da repetição incompleta (Olaisen *et al.*, 1998).

Sempre que seja desenvolvido um novo marcador genético o autor deverá submetê-lo à Comissão de DNA fornecendo detalhes como as condições de PCR, sequência dos *primers*, condições de visualização e sugestões de nomenclatura. O sistema a validar deverá ser publicado num jornal forense (Bar *et al.*, 1992).

Em 2002 foram publicadas as primeiras recomendações do grupo de trabalho do cromossoma Y do ISFG (*International Society of Forensic Genetics*) que ajudam a uniformizar e clarificar problemas relacionados com a nomenclatura, definição dos *loci* e classificação dos alelos, descrição de metodologias e aplicações na genética forense e de populações. Essas recomendações salientam algumas regras que permitiram uniformizar a nomenclatura para os vários *loci* descritos no cromossoma Y nomeadamente em relação à designação que deve ser feita pelo sistema de D#S# (Gill *et al.*, 2001).

A maioria dos microsatélites do cromossoma Y é constituída por repetições tetranucleótídicas que consistem em sequências variantes e não variantes. A nomenclatura destes *loci* é baseada no número total de repetições quer sejam variáveis ou não. Se uma nomenclatura já estiver em vigor deverá continuar a ser utilizada mas para novos *loci* é recomendado que os alelos sejam designados segundo o número de repetições, sejam elas variáveis ou não (Olaisen *et al.*, 1998). Em casos mais complicados a nomenclatura a utilizar deverá ser segundo a primeira publicação desse *loci*. Em sistemas com duplicação como o DYS385 os alelos devem ser separados por um hífen (Olaisen *et al.*, 1998).

1.2 APLICAÇÕES DOS MICROSATÉLITES

Os microsátélites são óptimos marcadores para a análise genética e tornaram-se importantes na realização de estudos populacionais (Di Rienzo *et al.*, 1998). A importância funcional destes marcadores é ainda desconhecida apesar de algumas vezes aparecerem em exões associados a algumas doenças. Expansões e contracções de microsátélites foram já associadas a alguns tipos de cancros e doenças hereditárias (Kunkel, 1993) estando normalmente os alelos grandes fortemente associados a patologias de origem genética (Sutherland e Richards, 1995).

Há evidências que as sequências de microsátélites também têm um papel funcional como codificadores ou elementos de regulação (Kunzler *et al.*, 1995; Kashi *et al.*, 1997). As regiões promotoras que contêm microsátélites funcionam como elementos iniciadores de expressão e a existência de deleções nesses microsátélites podem reduzir a sua actividade (Moxon *et al.*, 1994). Os microsátélites que se encontram nas regiões promotoras dos genes podem influenciar a expressão desses genes mas nem todos os microsátélites presentes nas regiões promotoras estão associados a actividades reguladoras (Kashi e Soller, 1999).

Parece haver restrições em relação ao número possível de repetições nos microsátélites pois se assim não fosse este número poderia aumentar ou diminuir indefinidamente, o que não acontece. Há discrepâncias entre o número de repetições que se observa e o que seria de esperar num *locus* neutral, que podem ser justificadas por mecanismos como a conversão génica ou uma tendência no processo mutacional (há evidências de uma tendência maior para o aumento de repetições em detrimento da sua diminuição). A selecção parece ser também um factor de relevo uma vez que foram encontradas evidências de associação de microsátélites à transcrição de genes, nomeadamente repetições TG, que aparecem tanto em macacos como em humanos, sugerindo que estas repetições podem influenciar a expressão desses genes (Hamada *et al.*, 1984).

Os microsátélites podem ser utilizados para identificar associações de vários *loci* a doenças, permitindo a determinação da influência de um segundo *locus* na susceptibilidade a uma doença, se existir entre o segundo *locus* um desequilíbrio positivo

ou negativo em relação ao primeiro *locus* que possa influenciar a predisposição a essa doença (Carrington *et al.*, 1999).

Os microsatélites são altamente polimórficos e têm sido por isso utilizados em estudos de populações humanas (Bowcock *et al.*, 1994) estudos de parentesco (Morin *et al.*, 1994) e forenses (Jeffreys *et al.*, 1992) nomeadamente na identificação de corpos em desastres de massa (Clayton *et al.*, 1995). Podem ser ainda utilizados para definir distâncias genéticas entre populações (Rubinsztein *et al.*, 1995a; Goldstein e Pollock, 1997) bem como em estudos de taxonomia molecular, evolução e genética de populações (Bowcock *et al.*, 1994) e no mapeamento genético (Weissenbach *et al.*, 1992).

1.2.1 Estudos populacionais

O objectivo fundamental da genética de populações é perceber a frequência dos alelos de vários genes evidenciando o poder da selecção positiva pois os genes que apresentam mais ou menos polimorfismos do que o esperado na base do seu nível de divergência, são candidatos a selecção positiva (Payseur *et al.*, 2002).

A selecção natural e o desequilíbrio numa população podem levar a desvios na distribuição das frequências dos polimorfismos mas tratando-se de um desequilíbrio na população este afectará todos os *loci* de modo igual enquanto a selecção afectará apenas algumas regiões do genoma (Payseur *et al.*, 2002).

O excesso de alelos raros indica a manutenção da difusão de alelos ligeiramente prejudiciais mas também pode representar uma evidência de uma expansão recente de uma população. Outra explicação é que o modelo SMM é inadequado para muitos dos microsatélites nos humanos e que se está sempre a prever um elevado valor da heterozigotia esperada (Payseur *et al.*, 2002).

Quando uma população está sujeita a uma redução efectiva do seu tamanho, observa-se um excesso de heterozigotia em alguns *loci* neutrais. Este excesso de heterozigotia persiste só durante algumas gerações até ser restabelecido um novo equilíbrio (Cornuet e Luikart, 1996). Para os *loci* neutrais o número de alelos e a distribuição das suas frequências em populações naturais resultam de um equilíbrio entre a ocorrência de

mutações e a deriva genética. Quando ocorre um efeito de gargalo na população (*bottleneck*) e o número efectivo da população diminui observa-se uma correlação positiva entre o número de alelos e a heterozigotia. Para testar se existe uma deficiência de alelos é necessário determinar a relação que existe entre a heterozigotia observada e a heterozigotia esperada, a partir do número de alelos esperado. Populações que apresentem um excesso de heterozigotia devem ser consideradas como tendo estado expostas a um efeito de gargalo recente. Um défice ou um excesso de heterozigotia pode ocorrer depois de uma recente alteração no número efectivo da população mas também pode representar uma desvantagem ou vantagem selectiva (Cornuet e Luikart, 1996). O mesmo se aplica se em vez da heterozigotia considerarmos a homozigotia (Watterson, 1978). No entanto, quando se trata de selecção, apenas alguns *loci* serão afectados enquanto após um efeito gargalo observa-se idêntico comportamento para todos os *loci* estudados, logo um défice de heterozigotia em vários *loci* poderá indicar uma expansão da população. Emigrações recentes podem contribuir para alguma confusão principalmente se os imigrantes vierem de populações geneticamente divergentes visto que esses imigrantes poderão aumentar rapidamente o número de alelos raros na população sem aumentar significativamente a heterozigotia falsificando um aumento ou escondendo uma diminuição do tamanho da população (Cornuet e Luikart, 1996).

Os mini e os microsatélites são os melhores *loci* para detectar efeitos de gargalo numa população, especialmente os microsatélites com interrupções ou microsatélites compostos por motivos de diferentes tamanhos, devido à sua elevada variabilidade (Estoup *et al.*, 1995).

Sob o modelo de alelos infinito, a configuração alélica de uma população num estado constante é especificado pelo tamanho da amostra e o número de alelos observado (Ewens, 1972). Este pressuposto conduz ao teste da neutralidade onde a homozigotia é determinada através do número de alelos observados (Watterson, 1978). Este valor de homozigotia esperado é então comparado com o valor observado e as discrepâncias indicam uma acção potencial da selecção positiva, apesar de esses desvios também poderem ser causados por fenómenos de expansão ou contracção das populações (Ewens, 1972). Dois modelos de selecção positiva, direcciona ou equilibrada podem causar impactos na diversidade molecular de diferentes maneiras: a direcciona pode

levar a uma redução da variabilidade e a um excesso de alelos; a equilibrada pode levar ao aumento de ligação de polimorfismos e a um déficit de alelos (Payseur *et al.*, 2002).

Atendendo a que a selecção pode levar a desvios na distribuição das frequências em *loci* que se encontram ligados com um conjunto de *loci* do genoma, poder-se-á prever quais as regiões do genoma que sofreram recentemente selecção positiva. Quando se observa a existência de um excesso de alelos raros poderá dizer-se que se trata de selecção direccionada enquanto um excesso na frequência de alguns alelos intermédios se deverá a selecção equilibrada (Payseur *et al.*, 2002).

Diferenças no genótipo de *loci* situados em zonas não codificantes do DNA, que são usadas como marcadores moleculares, como os microsatélites, podem não contribuir directamente para a *fitness* mas a correlação entre a heterozigotia em *loci* neutrais e a *fitness* de um indivíduo pode dever-se ao facto destes marcadores se encontrarem directamente ligadas aos *loci* que afectam a *fitness* ou então, a heterozigotia destes marcadores pode estar correlacionada com a *fitness* porque reflecte a heterozigotia do genoma. A diminuição da heterozigotia em *loci* neutrais é menor do que a esperada pela teoria havendo estudos em que se sugere que o poder da selecção contra a consanguinidade (mais homozigóticos) pode ser muito mais forte do que o previsto (Pemberton *et al.*, 1999).

As espécies que têm populações maiores deverão ter microsatélites mais longos pois quanto maior for a população, maior a heterozigotia e maior a probabilidade de ocorrência de mutações devido aos heteroduplexes (Amos, 1996; Rubinsztein *et al.*, 1995b).

As mutações nas células germinativas ocorrem mais frequentemente em indivíduos heterozigóticos em que os alelos difiram muito no número de repetições. Neste caso os microsatélites devem assegurar um correcto alinhamento dos cromossomas homólogos durante a meiose e diferenças grandes no número de repetições vão desestabilizar o emparelhamento dos cromossomas levando a recombinações anormais nas regiões flanqueadoras dos microsatélites (Pardue *et al.*, 1987). A taxa de mutação deve ser relacionada com a heterozigotia e é de esperar que os *loci* em populações grandes se desenvolvam mais depressa do que em populações pequenas (Amos, 1999). O tamanho e a estrutura das populações podem consequentemente afectar a taxa de mutação dos

microsatélites uma vez que elas são o resultado de recombinação devido a heterozigotia (Rubinsztein *et al.*, 1999).

Se a miscigenação das populações actua na aceleração de expansões é provável que a incidência de doenças devido a repetições de tripletos seja elevada e aumente mais rapidamente em populações onde os eventos históricos juntam pessoas de diferentes origens (Rubinsztein *et al.*, 1999).

Podem-se determinar fenómenos de expansão de populações observando a variabilidade genética dos microsatélites através da utilização de parâmetros como a heterozigotia e variação no tamanho dos alelos (Kimmel *et al.*, 1998).

Conhecer a estrutura genética de uma população e poder fazer inferências acerca da sua ancestralidade é importante, porque algumas doenças frequentes e a resposta a drogas estão muitas vezes associados a variantes genéticas cuja frequência difere entre grupos raciais (Bamshad, 2005). Regiões do genoma que apresentam grandes diferenças entre populações são regiões candidatas de terem sofrido selecção local (Schlötterer, 2002) logo um estudo do genoma é uma maneira eficaz de identificar essas regiões (Kayser *et al.*, 2003).

1.2.2 Testes de paternidade e identificação de indivíduos

A identificação de indivíduos é necessária em muitos casos nomeadamente em circunstâncias forenses. Desde 1985 que se têm utilizado metodologias de tipagem de DNA que permitiram resolver casos de identificação de indivíduos, utilizando baterias de RFLP-VNTR muito polimórficas, (Jeffreys *et al.*, 1985). Hammond e colaboradores (1994) validaram um método baseado na PCR utilizando polimorfismos comuns de DNA que permitiram uma rápida e fiável identificação de indivíduos. Foi desenvolvida uma bateria de microsatélites com um grande poder de discriminação utilizando sequências repetitivas de três, quatro e cinco pares de bases que permitem a identificação de indivíduos e a determinação de parentescos. Para a validação desses microsatélites utilizaram várias populações e estudaram parâmetros como o equilíbrio de *Hardy-Weinberg* apresentando estatísticas que reforçam o seu potencial para a identificação de indivíduos tendo também determinado a taxa de mutação para cada *locus*. Alford e

colaboradores (1994) validaram também a mesma metodologia utilizando para isso trios constituídos pela mãe, filho e pressuposto pai.

Em casos de investigações de paternidade é normalmente aceite que haja pelo menos 2 exclusões para concluir a exclusão da paternidade, havendo no entanto pelo menos um caso reportado em que apesar de se observar duas exclusões se veio a confirmar a paternidade sendo pois recomendado que o critério seja de pelo menos 3 exclusões (Gunn *et al.*, 1997).

A utilização de microsatélites do cromossoma Y na identificação de indivíduos e até em testes de paternidade é sugerido apenas para confirmar as exclusões pois o mesmo haplótipo pode ser partilhado por indivíduos aparentados por via paterna (Kayser *et al.*, 1997). A maior parte das bases de dados do cromossoma Y apresenta frequências de haplótipos para grandes regiões geográficas, populações que partilham a mesma linguagem ou que habitam nas mesmas zonas geográficas. No entanto estes resultados só serão válidos se não existirem subestruturas populacionais pois estas têm sido observadas em alguns grupos dentro das mesmas populações (Beleza *et al.*, 2006).

Os microsatélites tornaram-se marcadores excelentes para estudos com DNA antigo. O entendimento de evolução dos comportamentos sociais depende da possibilidade de diferenciar geneticamente indivíduos e conseguir estimar com suficiente precisão a relação genética entre pares ou grupos de indivíduos (Queller e Goodnight, 1989). Por exemplo para estudar as variações no gene da lactose utilizou-se DNA antigo extraído a partir de vários esqueletos e nesse estudo foram utilizados 10 microsatélites, para excluir a possibilidade de contaminação dos vários esqueletos (Burger *et al.*, 2007).

1.2.3 Associação com doenças

Nos últimos anos têm sido utilizados métodos para encontrar mutações associadas a doenças, baseados em estudos familiares que envolvem análises de *Linkage* mas que se têm revelado de baixa resolução (Rannala e Reeve, 2001). Mesmo em situações em que se estudam centenas de famílias o método não tem permitido estabelecer relações quando as distâncias no genoma são maiores do que 1cM (Boehnke, 1994). Por outro

lado estudos de LD parecem ser vantajosos pelo facto de recorrerem a uma amostra grande de cromossomas de indivíduos não aparentados (Bodmer, 1986). As mutações que provocam as doenças aparecem associadas a haplótipos constituídos por vários marcadores nos cromossomas. Esses haplótipos e as doenças estarão inicialmente associados numa determinada população. Passado algum tempo essa associação diminui devido à recombinação, com uma taxa determinada pela distância genética dos marcadores à mutação que provoca a doença. Uma das vantagens de mapeamento por LD de fina escala, na era pós-genoma, é que este método parece ser mais robusto para o mapeamento de genes relevantes no estudo das doenças genéticas complexas que são poligénicas e influenciadas pelo ambiente. Foi recentemente sugerido que os genes ligados a essas doenças podem ser divididos em duas classes, uma constituída por genes com um efeito marginal e alta frequência e outro por genes com alta influência e baixa frequência (Risch, 2000).

1.2.3.1 Expansão de trinucleótidos

Um interesse adicional em relação aos microsatélites surge com a descoberta da existência de uma relação entre o número de repetições trinucleotídicas e nove doenças nos humanos associadas a 11 *loci*. As repetições são normalmente CGG/CCG ou CAG/CTG (Ashley e Warren, 1995) e são também designadas por mutações dinâmicas (Richards e Sutherland, 1997).

Estas repetições possuem a capacidade de formar ganchos constituindo uma estrutura que se encontra no limiar da estabilidade, sendo esta influenciada pela sequência flanqueadora do gene. A estabilidade desses ganchos depende de dois factores, a sequência e o comprimento do DNA pois só algumas sequências com determinado comprimento podem formar ganchos estáveis, existindo uma correlação entre a capacidade de formar ganchos, a sequência de expansão e o comprimento dessas expansões. A formação de ganchos disponibiliza uma estrutura base para a estabilidade da região CCG do gene da Doença de Huntington e explica o efeito estabilizador das interrupções AGG no gene FMR1 (Gacy *et al.*, 1995).

A sequência da repetição é importante pois repetições CTG formam estruturas em gancho devido à metilação mas o mesmo não acontece se as repetições forem ATC ou GAT (Mitas *et al.*, 1995).

Quando as repetições estão localizadas no gene ou perto das sequências transcritas podem ter um efeito na transcrição do gene que se pode manifestar na forma de doença. Para cada *locus* em que ocorrem as mutações dinâmicas há um número normal de cópias acima do qual as repetições se tornam instáveis, isto é, o número de cópias pode aumentar até que haja manifestações na forma de doença ou de locais frágeis (Richards e Sutherland, 1997).

Muitas doenças associadas a repetições trinucleotídicas CGG e AGC apresentam fenótipos diferentes consoante o número de repetições, nomeadamente ao nível da idade do aparecimento da doença e da severidade dos sintomas estando as expansões normalmente associadas à ausência de transcrição e consequentemente à perda da função (Richards e Sutherland, 1997).

Em 1991 é descrito o síndrome do X-frágil como a primeira doença genética associada a mutações dinâmicas de microsatélites (Fu, 1991; Kremer *et al.*, 1991; Oberlé *et al.*, 1991). Esta doença é causada por uma mutação dinâmica da repetição trinucleotídica CCG na região do gene *FMR1* (Verkerk *et al.*, 1991) sendo a variação do número de cópias a base da instabilidade deste *locus* (Kremer *et al.*, 1991).

Muitas doenças neurológicas têm sido associadas a mutações dinâmicas de repetições trinucleotídicas como por exemplo a atrofia muscular espino-bulbar associada a repetições do tipo AGC (La Spada *et al.*, 1991) e a distrofia miotónica que surge associada a expansões tanto de repetições AGC (Brook *et al.*, 1992) como de repetições CTG (Imbert *et al.*, 1993). O agravamento da severidade desta doença ao longo de gerações é justificado pela existência de mutações dinâmicas que fazem aumentar o número de repetições, alterando assim os fenótipos da doença dentro da mesma família (Harley *et al.*, 1993).

Foram também encontradas pequenas expansões de trinucleótidos junto aos genes que codificam poli-glutaminas, associadas a doenças neurodegenerativas (Sutherland e Richards, 1995). Neste caso encontra-se a doença de Huntington, provocada pela expansão e instabilidade da sequência repetitiva CAG no gene *IT15* no cromossoma 4

(MacMillan *et al.*, 1993), apresentando um número de repetições entre 9 e 34 em indivíduos normais e variando entre 30 a 86 repetições em indivíduos com a doença (Duyao *et al.*, 1993; Snell *et al.*, 1993). Neste caso o número de repetições está inversamente correlacionado com a idade de início da doença, com uma transmissão meiótica instável em mais de 80% e maior incidência nas transmissões paternas (Duyao *et al.*, 1993).

A doença de Machado-Joseph (MJD) também está associada a repetições CAG situadas no cromossoma 14 (locus 14q32.1). Em indivíduos normais o gene contém entre 13 e 36 repetições CAG enquanto os doentes apresentam expansões que variam entre 68 e 79 repetições (Kawaguchi *et al.*, 1994).

Existem também repetições dinucleotídicas CG (*CpG islands*) que são locais críticos para a ocorrência de mutações, estando cerca de 75% dos dinucleótidos CpG metilados pela enzima metiltransferase, que adiciona à sequência um grupo 5-metilcitosina (Jobling *et al.*, 2004). A metilação tem um papel importante no controlo da expressão dos genes e no controle da condensação dos cromossomas (Jobling *et al.*, 2004). A associação destas *CpG islands* a repetições de trinucleótidos leva a crer que as regiões flanqueadoras e a estrutura da cromatina perto dos tripletos desempenham um papel importante na instabilidade das repetições (Gourdon *et al.*, 1997).

As repetições de trinucleótidos são frequentes no genoma humano podendo ser encontradas no *Genebank* mais de 100 *loci* humanos com repetições AGC/GTC com mais de 5 cópias (Caskey *et al.*, 1992). É por isso previsível o aumento de doenças associadas a este tipo de repetições.

1.2.3.2 Perda de heterozigotia e instabilidade de microsatélites

A perda de heterozigotia (LOH - *Loss of Heterozigosity*) está frequentemente associada a determinados tumores nos humanos nomeadamente o cancro colo rectal (Halling *et al.*, 1999), da mama (Schneider-Stock *et al.*, 1998), da próstata (Werely *et al.*, 1996), da bexiga (Aguirell *et al.*, 1992) dos ossos (Bovée *et al.*, 1999) e leucemias (Pakkala *et al.*, 1991).

No DNA dos tumores do colo rectal observa-se a existência de instabilidade de repetições (CA)_n nos cromossomas 5q, 15q, 17p e 18q, nas células somáticas. São observadas diferenças entre o DNA dos tecidos normais e o dos tecidos tumorais em cerca de 28% dos casos. Esta instabilidade dos microsatélites está significativamente correlacionada com a localização do tumor e com a sobrevivência dos pacientes e inversamente com a perda de heterozigotia nos cromossomas 5q, 17p e 18q, o que sugere que o mecanismo de aparecimento do cancro colo rectal poderá não envolver directamente a perda de heterozigotia (Thibodeau *et al.*, 1993).

Por comparação do tecido tumoral e de um outro tecido do indivíduo, observam-se alterações que permitem a detecção precoce (Serra *et al.*, 2007) assim como a determinação da sua forma de aparecimento e progressão, consoante estejam ou não associadas a LOH (Bovée *et al.*, 1999).

Designam-se por mutações RER (*Replication Error*) as mutações associadas a sequências repetitivas de DNA, que ocorrem essencialmente ao nível das células somáticas (Liu *et al.*, 1995) estando estas associadas a vários tipos de cancro nomeadamente ao cancro da mama (Schwarzenbach *et al.*, 2007), da próstata (Müller *et al.*, 2008) e ao carcinoma colo rectal (Hoff-Olsen *et al.*, 1998). A maioria dos cancros de cólon hereditários é caracterizada por instabilidade de microsatélites, isto é, a acumulação de mutações somáticas no tecido tumoral devido a defeitos no mecanismo de reparação durante o emparelhamento (Di Rienzo *et al.*, 1998), enquanto nos cancros de cólon esporádicos a instabilidade de microsatélites aparece em apenas cerca de 11 a 28% dos casos e só 43% desses têm mutações nos genes de reparação do DNA (Honchel *et al.*, 1995).

Indivíduos com mecanismos de reparação deficiente, normalmente devido a mutações nos genes *MSH2* ou *MLH1* (Thibodeau *et al.*, 1996; 1998), têm uma elevada frequência de cancro do colo rectal associados ao aparecimento de instabilidade de microsatélites nas células somáticas (Armour, 2006). Mutações somáticas de microsatélites em células cancerosas podem resultar da falta de fidelidade durante a replicação, tendo estas células menos capacidade para corrigir mutações apresentando uma elevada taxa de mutações (Di Rienzo *et al.*, 1998).

1.2.3.3 Variação do tamanho dos alelos no Receptor dos Androgéneos

Os Androgéneos e os receptores dos Androgéneos (AR) são essenciais para o desenvolvimento e manutenção do fenótipo masculino e da espermatogénese. O gene AR está localizado na região q11-12 do cromossoma X (Lubahn *et al.*, 1988) e é constituído por 8 exões (Ferlin *et al.*, 2004). O receptor exibe dois *loci* polimórficos localizados no exão 1, caracterizados por diferentes números de repetição CAG e GGC resultando em diferentes tamanhos de poli-glutaminas e poli-glicinas no N-terminal TAD da proteína AR, que parece controlar a função AR. O número de repetições CAG varia entre 10 a 35 (média 21 a 23) enquanto o de GGC varia entre 4 a 24 (média 16 a 17), em homens normais. Mutações no gene AR causam uma variedade de defeitos relacionados com a insensibilidade aos Androgéneos, nomeadamente infertilidade (Aiman *et al.*, 1979; Hiort *et al.*, 2000). Longas repetições CAG resultam numa actividade transcricional reduzida (Chamberlain *et al.*, 1994, Choong *et al.*, 1996). Por outro lado, pequenas repetições CAG representam pequenas extensões de poli-glutaminas que provocam uma maior transcrição de AR, associado ao aumento do risco de cancro de próstata (Giovannucci *et al.*, 1997; Hakimi *et al.*, 1997; Stanford *et al.*, 1997). Aumentos moderados nas expansões de poli-glutaminas, apesar de estarem no intervalo normal de variações, podem também alterar a função de AR. Existem estudos que descrevem a existência de uma associação entre o número de repetições CAG e GGC e a infertilidade masculina e apesar de muitas vezes não serem encontradas diferenças significativas na distribuição de CAG e GGC quando consideradas em separado, quando se juntam aparecem haplótipos que apresentam diferenças significativas entre pacientes e controlos. Alguns desses haplótipos estão associados ao aumento da susceptibilidade à infertilidade como CAG=21/GGC=18 e CAG≥21/GGC≥18 com um risco relativo de 2,47 e 1,6 respectivamente enquanto os haplótipos CAG≥23/GGC≤16 apresentam um risco relativo de 0,09 parecendo conferir protecção contra a infertilidade (Ferlin *et al.*, 2004).

Repetições curtas de GGC parecem aumentar o risco de cancro da próstata (Hakimi *et al.*, 1997; Stanford *et al.*, 1997; Chang *et al.*, 2002). Alelos longos de CAG e GGC estão relacionados com a diminuição da actividade do receptor e associados ao cancro da mama (Suter *et al.*, 2003) enquanto alelos de tamanhos pequenos de CAG<19 e GGC<15

resultam num aumento da actividade do receptor e estão associados ao cancro da próstata, a idade precoce do aparecimento da doença e um grau de desenvolvimento maior na altura do diagnóstico (Chen *et al.*, 2002;).

Em ambos os microsatélites foram observadas diferenças significativas nos tamanhos dos alelos entre populações africanas e não africanas sendo o número de repetições CAG menor nos Afro-americanos, intermédia nos europeus e mais alta nos asiáticos (Ferlin *et al.*, 2004). Este facto pode ser importante para justificar as diferenças étnicas observadas na incidência de determinados cancros como o cancro da próstata que apresenta uma incidência maior nos afro-americanos do que nos americanos descendentes de europeus sendo muito baixa nos asiáticos (Hsing *et al.*, 2000). Em casos de estudos de associação é muito importante considerar a possibilidade de existência de estratificação da população pois as diferenças na frequência dos alelos podem levar a confusões nos resultados de associações por se deverem apenas a uma subestruturação da população em estudo.

Os estrogéneos estão associados à regulação de vários aspectos do comportamento sendo algumas alterações atribuídas aos polimorfismos no gene receptor dos androgéneos (Westberg *et al.*, 2003). Várias experiências demonstraram que dois tipos diferentes de células neurais podem ser afectadas em consequência de mutações no gene *AR* podendo por isso considerar-se este gene associado a alterações neurológicas (Caskey *et al.*, 1992). Foi também encontrada uma associação significativa entre o baixo número de repetições GGC e o comportamento agressivo e impulsividade sexual em ambos os sexos (Comings *et al.*, 2002).

Vários estudos apontam para o efeito protector do baixo número de repetições CAG no *AR* em relação ao cancro cerebral e do pescoço (Santos *et al.*, 2004) mas no entanto contribuem para um maior risco do cancro da mama (Elhaj *et al.*, 2001), enquanto o número elevado de repetições GGC parece conferir um efeito protector (Suter *et al.*, 2003).

1.3 HISTÓRIA DAS POPULAÇÕES

Foram estudadas as populações dos arquipélagos da Madeira, Açores e Cabo Verde localizados no Oceano Atlântico (figura 3). Foram estudadas ainda a população de Portugal continental atendendo a que estes arquipélagos foram todos colonizados pelos portugueses no século XV e a população da Guiné-Bissau, uma das principais origens dos escravos durante a colonização das ilhas atlânticas.

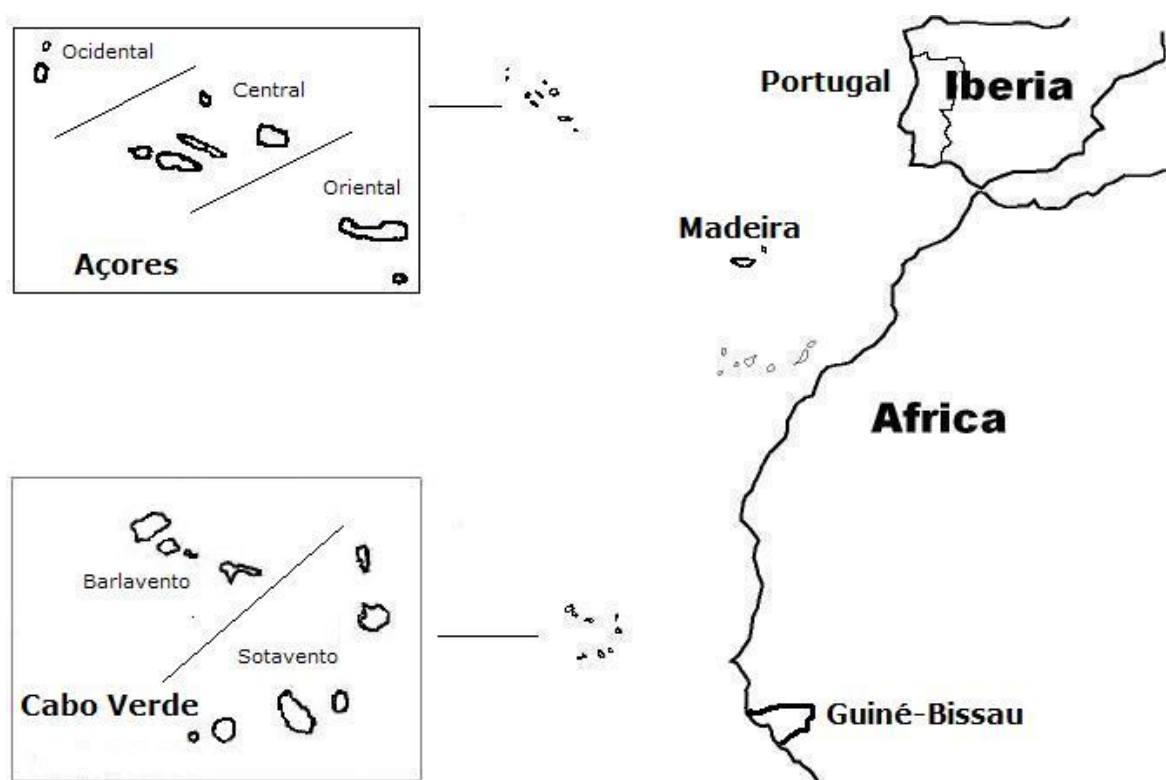


Figura 3: Mapa de localização dos arquipélagos da Madeira, Açores e Cabo Verde situados no Oceano Atlântico.

1.3.1 Arquipélago da Madeira

O Arquipélago da Madeira é constituído por duas ilhas habitadas, Madeira e Porto Santo com cerca de 250.000 habitantes. Foi descoberto pelos portugueses em 1419,

tendo sido iniciado o seu povoamento entre 1420 e 1425. Entre 1440 e 1450 é feita a doação das terras aos capitães donatários (João Gonçalves Zarco, Tristão Vaz Teixeira e Bartolomeu Perestrelo) que lideraram o processo de colonização, ficando o arquipélago dividido em três capitanias: Funchal, Machico e Porto Santo (Pereira, 1989).

O povoamento começou na costa sul, com nobres portugueses, italianos, espanhóis e, associados à cultura do açúcar vieram também escravos do Norte de África (Russell-Wood, 1998).

A partir daí dá-se um grande crescimento na população, passando de 150 famílias em 1440, para cerca de 3000 habitantes em 1460, aumentando para cerca de 15000 habitantes no início do século XVI. Durante esse período cerca de 50% dos povoadores vieram do Norte de Portugal (Braga, Porto e Viana do Castelo) (Carita, 1999). Em finais do século XV e início do século XVI, devido ao comércio do açúcar, há entrada e fixação de italianos, flamengos, franceses e bretões que se fixam nas vertentes meridionais da ilha, surgindo como os principais proprietários em várias localidades nomeadamente João de Bettencourt na Ribeira Brava e João Esmeraldo na Ponta do Sol (Carita, 1999).

No início do século XVI havia cerca de 2000 escravos vindos da Índia (em Machico um dos capitães tinha 12 escravos indianos) e norte de África (Carita, 1999) e, depois da descoberta de Cabo Verde, a Madeira começou a receber escravos da costa Oeste Africana, que chegaram a constituir, em meados do século XVI, 10% da população (Pereira, 1989).

Os escravos dispersaram-se por toda a ilha estabelecendo-se em colónias mais numerosas nas regiões cultivadas de cana-de-açúcar como Faial, Machico, Santa Cruz e Ponta do Sol (Pereira, 1989).

A maior parte dos escravos e libertos existentes na ilha nos séculos XVI e XVII nasceram no Funchal nas freguesias da Sé e São Pedro. Há também registos de nascimentos no século XVI nas freguesias de Ribeira Brava e Caniço e no século XVII em Santa Cruz, Porto Santo, Ponta Delgada, Estreito da Calheta, Caniço e Câmara de Lobos. Neste período, dos 3.168 escravos baptizados, apenas 6% se apresentavam com pai e mãe, sendo todos os outros ilegítimos e muitos deles resultando de relações de concubinato entre as escravas e os senhores (Vieira, 1991).

A costa norte incluída na capitania de Tristão Vaz Teixeira manteve-se entre os séculos XV e XVIII vinculada às estruturas de poder sedeadas em Machico enquanto na costa sul vários municípios surgiram logo no início do século XVI como Ponta do Sol e Calheta (Carita, 1999). No Porto Santo ficaram como testemunho dos primeiros colonizadores até aos nossos dias nomes como Perestrelo, Baião, Brito, Castro, Coelho, Ferreira, Ruas e Vasconcelos (Pereira, 1989).

João Gonçalves Zarco deixou uma larga descendência que se espalhou pela Madeira, Açores, Brasil e continente português (Silva e Menezes, 1984). No século XIX cerca de 2000 madeirenses foram expulsos da Madeira por se terem convertido ao Protestantismo, por influência do Escocês Robert Reid Kalley, tendo-se espalhado até à ilha da Trindade, das Caraíbas para o Ilínois, nos Estados Unidos, e daí para o arquipélago do Havai (Fernandes, 2004).

1.3.2 Arquipélago dos Açores

É constituído por nove ilhas divididas em três grupos: Oriental (São Miguel e Santa Maria), Central (Terceira, São Jorge, Pico e Horta) e Ocidental (Flores e Corvo) com cerca de 240.000 habitantes sendo a ilha de São Miguel a mais povoada com cerca de 150.000 habitantes.

Foi descoberto pelos portugueses em 1431, sendo Santa Maria a primeira ilha a ser descoberta logo seguida de São Miguel, Terceira, Faial e Pico mas só começou oficialmente a ser povoado em 1439. Só em 1452 foram descobertas as ilhas do grupo Ocidental, Corvo e Flores. O povoamento do grupo Central foi fortemente influenciado pelos Flamengos, pois já em 1490 estavam instalados nestas ilhas cerca de 2000 indivíduos provenientes da Flandres (Rogers, 1979). A capitania da ilha Terceira é doada a Jácome de Bruges, da Flandres enquanto o Faial é colonizado por Jacob von Hurter da Nuremberga. A partir do Faial, os flamengos teriam irradiado para as ilhas do Pico, São Jorge e mais tarde para as Flores que, no final do século XV, se chamavam ilhas flamengas (Frutuoso, 1977). Este arquipélago teve um papel importante no comércio marítimo entre a América e a Europa durante os séculos XV e XVI. A sua posição estratégica atraiu

povoadores de diferentes origens, a maior parte portugueses mas também judeus, escravos africanos, flamengos, franceses, italianos, ingleses e espanhóis (Guill, 1993).

1.3.3 Arquipélago de Cabo Verde

É constituído por 9 ilhas divididas em 2 grupos: Sotavento (Santo Antão, São Vicente, São Nicolau) e Barlavento (Santiago, Brava, Fogo, Boavista, Sal e Maio) sendo esta última a única ilha desabitada. O conjunto das ilhas tem cerca de 463.000 habitantes, sendo a de Santiago a mais povoada, com cerca de 240.000 habitantes.

O arquipélago foi descoberto em 1460 pelos portugueses e a sua colonização começou pela ilha de Santiago que foi dividida em duas capitánias, tendo como donatários o italiano António de Noli e Diogo Afonso, com privilégios iguais aos que foram concedidos para a Madeira e Açores (Serrão, 1980). Em 1461 são descobertas as ilhas de São Nicolau, Santa Luzia, Santo Antão e São Vicente. Com o avanço meridional da expansão e, mais tarde, com a abertura da rota da Índia por via do Brasil, o arquipélago passou a desempenhar um papel fundamental para a navegação atlântica (Serrão, 1980), servindo durante o período da escravatura como um interposto no comércio transatlântico de escravos (Russell-Wood, 1998).

Os escravos eram capturados na região da Senegâmbia (entre o Senegal e a Guiné) e trazidos para as ilhas de Cabo Verde. A colonização envolveu alguns homens europeus (principalmente ibéricos) e escravos. Os europeus estabeleceram relações com as mulheres escravas originando os mulatos que constituem a população actual destas ilhas. A colonização foi feita em várias fases sendo Santiago a primeira ilha a ser colonizada seguida pelo Fogo e Brava. Só no século XVI e especialmente durante o século XVII é que foram colonizadas as outras ilhas (Barcellos, 1899).

1.4 OBJECTIVOS

O objectivo geral deste trabalho é o estudo das aplicações dos microsátélites em várias vertentes, nomeadamente estudos populacionais, datação de mutações e diagnóstico de doenças.

Pretende-se com este trabalho:

Estudar as populações dos Arquipélagos da Madeira, Açores e Cabo Verde, situados no Oceano Atlântico e que foram colonizados pelos Portugueses no século XV, de modo a verificar a influência dessa colonização na população actual destes arquipélagos bem como a influência africana devido ao comércio de escravos no início da colonização. Para tal, recorrer-se-á ao estudo dos microsátélites neutrais autossómicos e do cromossoma Y.

Criar bases de dados das 3 populações que possam ser utilizadas em averiguações de paternidade, onde a frequência dos alelos é importante para a determinação do Índice de Paternidade e consequentemente da Probabilidade de Paternidade.

Definir haplótipos, associá-los aos haplogrupos (que são definidos por mutações pontuais) e estudar a influência das linhagens paternas nos três arquipélagos, utilizando microsátélites do cromossoma Y e comparar com estudos de DNA mitocondrial para as mesmas populações, o que permitirá analisar as diferenças que possam existir na colonização em relação às linhagens masculina e feminina.

Averiguar a existência de fenómenos como o efeito fundador e deriva genética nas populações da Madeira e dos Açores, bem como a possível ocorrência de mutações pontuais recorrentes no cromossoma Y, utilizando a variabilidade dentro dos microsátélites do cromossoma Y e a construção de diagramas em rede (*networks*).

Definir a idade de coalescência de mutações, nomeadamente as detectadas no gene *DAZ* (*Delected in Azospermia*) localizado na região *AZFc* (*Azoospermic Factor*) do cromossoma Y, recorrendo aos microsátélites do cromossoma Y.

Estudar o número de repetições GGC no gene *FMR1* (*Fragile X Mental Retardation*), situado no cromossoma X e associado à síndrome do X-Frágil, nas populações da Madeira

e da Guiné-Bissau, o que permitirá inferir acerca da probabilidade da ocorrência de indivíduos doentes e pré-mutados nestas populações.

Determinar a frequência do número de repetições dos tripletos CAG e GGC do gene AR (*Androgen Receptor*) do cromossoma X em indivíduos com infertilidade masculina, provenientes de uma população do Norte de Portugal e em indivíduos que apresentam níveis de PSA elevados na população da Madeira, definir haplótipos e comparar as populações.

Analisar a variabilidade de dois polimorfismos do gene *IL4* (IL4RP2 e IL4590) localizados no cromossoma 5 e averiguar a existência de Desequilíbrio de Ligação (LD - *Linkage Disequilibrium*) entre dois microsatélites (D5S818 e CSF1PO) e estes dois polimorfismos. Será então possível inferir sobre a sua ancestralidade e evolução, de modo a perceber a diferença existente na frequência desses polimorfismos nas populações da Madeira, Açores, Cabo Verde, Guiné-Bissau e Portugal Continental.

2. MATERIAL E MÉTODOS

2.1 SELECÇÃO DAS AMOSTRAS

As amostras foram todas recolhidas em regime de voluntariado, e com consentimento informado. Os locais de colheita foram quartéis militares ou hospitais, mediante protocolos oficiais assinados entre a Universidade da Madeira e os governos ou instituições participantes. Os estudos envolvendo doentes tiveram a aprovação das comissões de ética das instituições participantes.

2.1.1. Estudos Populacionais

Foram seleccionadas amostras de sangue de indivíduos com base na origem familiar e pertencendo à mesma região há pelo menos três gerações, oriundos da Madeira, Açores e Cabo Verde, bem como de duas populações de referência, Portugal Continental, de onde provieram os primeiros colonizadores e Guiné-Bissau, que foi uma das principais origens dos escravos durante a colonização das ilhas atlânticas.

População da Madeira

O perfil de 106 indivíduos do arquipélago da Madeira foi determinado através do estudo dos microsatélites autossómicos (Publicação I: Fernandes *et al.*, 2002). Os indivíduos foram divididos em quatro áreas geográficas: Nordeste (n=49), Sudoeste (n=37), Funchal (n=13) e Porto Santo (n=7).

No estudo dos microsatélites do cromossoma Y foi determinado o perfil de 142 indivíduos do Arquipélago da Madeira (Publicação II: Fernandes *et al.*, 2001), divididos

também nas mesmas quatro áreas geográficas: Nordeste (n=47), Sudoeste (n=37), Funchal (n=35) e Porto Santo (n=23).

População dos Açores

Foram estudados 100 indivíduos do Arquipélago dos Açores, utilizando 16 microsatélites autossômicos (Publicação III: Velosa *et al.*, 2002) divididos em três grupos de ilhas: Oriental - S. Miguel (n=42) e Santa Maria (n=3); Central - Terceira (n=22), Faial (n=9), São Jorge (n=7) e Pico (n=10); Ocidental - Graciosa, Flores e Corvo (n=6). Foram ainda estudados os microsatélites do cromossoma Y nos mesmos indivíduos (Publicação IV: Fernandes e Brehm, 2003).

População de Cabo Verde

Foram definidos os perfis de 161 indivíduos de oito das nove ilhas, divididos em 2 grupos: Barlavento – São Nicolau (n=12), Santo Antão (n=55) e São Vicente (n=18) e Sotavento – Fogo (n=22), Santiago (n=42), Brava (n=8), Boavista (n=4) e Sal (n=3) (Publicação V: Fernandes *et al.*, 2003).

População de Portugal Continental

A população de Portugal Continental foi subdividida em 3 grupos: Norte (n=120), Centro (n=127) e Sul (n=151) num total de 398 indivíduos tendo sido estudados apenas 5 microsatélites (CD4, FES, TPOX, TH01 e VWA) nesta população (Publicação VI: Fernandes e Brehm, 2002).

População da Guiné-Bissau

Foram estudados 101 indivíduos da população da Guiné-Bissau, divididos em 10 grupos étnicos: Balanta (n=10), Bijagós (n=10), Cassanga (n=9), Felupe (n=11), Fula (n=10), Mandinga (n=11), Manjaco (n=10), Massonca (n=10), Nálu (n=11) e Sussú (n=9) (Publicação VII: Gonçalves *et al.*, 2002).

2.1.2. Estudo da variabilidade de haplótipos no cromossoma Y

Foram definidos os haplótipos utilizando 11 microsatélites (DYS19, DYS389 I e II, DYS390, DYS391, DYS392, DYS393, DYS385, DYS438, DYS439 e DYS460) de indivíduos pertencentes ao haplogrupo E3b das populações do Norte de Portugal (n=21), Centro de Portugal (n=15), Sul de Portugal (n=8), Cabo Verde Norte (n=27), Cabo Verde Sul (n=13) e Guiné-Bissau (n=17) (Publicação VIII: Fernandes *et al.*, 2008a).

Foram ainda definidos haplótipos, utilizando os mesmos 11 microsatélites, de indivíduos de vários haplogrupos das populações da Madeira (n=140) e Açores (n=100) (Publicação IX: Fernandes *et al.*, 2008b).

2.1.3. Estudo da variabilidade do gene *DAZ* do cromossoma Y

Foram utilizados 95 indivíduos inférteis e 90 férteis, com o perfil do gene *DAZ* previamente determinado. Estas determinações foram realizadas no Serviço de Genética da Faculdade de Medicina do Porto. Foi determinado o haplótipo para todos os indivíduos, utilizando 10 microsatélites do cromossoma Y. (Publicação X: Fernandes *et al.*, 2006).

2.1.4. Detecção de trissomias por PCR

Foram utilizadas 120 amostras de líquido amniótico (1mL), colhidas por amniocentese a mulheres grávidas com mais de 35 anos que foram à consulta de obstetrícia do Hospital Central do Funchal, para despiste de anomalias cromossómicas principalmente devido a idade materna.

2.1.5. Estudo do número de repetições CAG/GGC do gene AR, associadas à infertilidade masculina

Foram utilizados 95 indivíduos inférteis e 90 férteis, assim considerados por terem pelo menos um filho. Os indivíduos inférteis foram divididos e classificados segundo a patologia apresentada. Estas amostras foram cedidas pelo Serviço de Genética da Faculdade de Medicina do Porto.

2.1.6. Estudo do número de repetições CAG/GGC do gene AR, associado aos níveis de PSA

Foram utilizadas para este estudo caso/controle 199 indivíduos do sexo masculino, seleccionados aleatoriamente entre os utentes do Laboratório de Patologia Clínica do Hospital Central do Funchal. Os indivíduos foram divididos em dois grupos pelos valores de referência de PSA segundo os parâmetros de referência definidos (Freedland *et al.*, 2004). Os indivíduos do grupo com PSA elevado ($n=97$), apresentavam valores de $PSA \geq 2$ ng/mL quando a idade era inferior a 50 anos e ≥ 4 ng/mL quando a idade era igual ou superior a 50 anos. Do grupo dos controlos ($n=102$) fazem parte os indivíduos com valores de PSA inferiores aos já referidos.

2.1.7. Determinação do número de repetições GGC do gene *FMR1* (*Fragile X Mental Retardation*)

Foram seleccionadas aleatoriamente 115 indivíduos da Madeira e 104 indivíduos da Guiné-Bissau da amostragem existente no Laboratório de Genética Humana. Foram utilizadas amostras de sangue de 26 indivíduos da Madeira com elevado QI, definido de acordo com o teste das inteligências múltiplas de Gardner (Gardner, 1993) e 25 indivíduos da Madeira com suspeita de Síndrome de X-Frágil.

2.1.8. Estudo da variabilidade de dois polimorfismos do gene *IL4*

Foram utilizados os resultados da tipagem dos microsatélites CSF1PO e D5S818 dos indivíduos da Madeira (n=93), Açores (n=85), Cabo Verde Norte (n=60), Cabo Verde Sul (n=101) e Guiné (n=104) para determinar a variabilidade de dois polimorfismos do gene *IL4* (RP2 e 590 C/T), localizado no cromossoma 5. Foram considerados apenas os indivíduos com os genótipos 183/183 TT (n=27), 253/253 TT (n=8) e 253/253 CC (n=124) (Publicação XI: Berenguer *et al.*, *submitted*).

2.2. EXTRACÇÃO DE DNA

O DNA foi extraído pelo método do Chelex (Walsh *et al.*, 1991) adaptado por Laréu *et al.* (1994). Este método rápido e barato (Hoff-Olsen *et al.*, 1999), consiste na utilização de uma resina (Chelex) que se agrega aos diversos componentes celulares, excepto DNA, permitindo a sua extracção de forma rápida e com boa qualidade para amplificação por PCR (Vigilant, 1999).

Os métodos baseados em resinas como o Chelex, quando comparados com métodos orgânicos, revelam-se melhores visto serem mais rápidos, requererem menos passos de transição, não necessitarem de reagentes perigosos e o custo dos reagentes ser menor. É o método mais recomendado na preparação do DNA para amplificação a partir de pequenas quantidades de sangue (Vandenberg *et al.*, 1997).

2.3. QUANTIFICAÇÃO DO DNA

A quantificação do DNA foi feita por espectroscopia de absorção. Este método baseia-se no facto dos constituintes da molécula de DNA absorverem radiações ultravioletas, na ordem dos 260 nm. A quantificação das amostras foi feita com o espectrofotómetro GeneQuant II (Pharmacia Biotech), o qual indicou valores na ordem dos 40 a 100 ng por microlitro de solução.

2.4. AMPLIFICAÇÃO DO DNA POR PCR

A técnica de PCR (*Polymerase Chain Reaction*) foi descrita por Mullis e colaboradores (Mullis *et al.*, 1986) e tem por objectivo aumentar a quantidade de sequências específicas da cadeia de DNA. A base teórica da PCR assenta no conhecimento da sequência nucleotídica e da organização funcional do DNA (Kleppe *et al.*, 1971), simulando *in vitro* a cascata de reacções que se processa nas células.

2.4.1. Microsatélites autossómicos

Fez-se a determinação do perfil dos indivíduos com os microsatélites autossómicos D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, TH01, CSF1PO, D16S539, Penta D, VWA, D8S1179, TPOX, FGA utilizando o kit PowerPlex®16 System (*Promega*).

A amplificação dos fragmentos de DNA foi feita por PCR (*Polymerase Chain Reaction*) com condições variáveis consoante os marcadores a estudar.

A mistura de PCR foi a do *kit* utilizado, excepto para os microsatélites isolados (CD4 e FES), marcados ou não com fluorescência, em que se utilizaram tampão 25 mM (*Firepol*), MgCl₂ 25 mM (*FirePol*), nucleótidos (*dNTPs*) a 2 mM (*Promega*) e os *primers* a 10 pmol/μl (*MWG Biotech*) e 1U *Taq-polymerase* a 5U/μl (*FirePol Solis Biodyne*), para um volume total de 25 μl.

2.4.2. Detecção de trissomias por PCR

Foram utilizados 8 microsatélites dos cromossomas 13 (D13S631, D13S634), 18 (D18S535, MBPB), 21 (D21S11, D21S1414, D21S1212) e X (XHRP) bem como a amelogenina que permitiu determinar o sexo de todos os indivíduos (tabela 1). A mistura de PCR foi: tampão B 10x (*Solis BioDyne*), MgCl₂ 25 mM (*Solis BioDyne*), *primers* 10 pmol/μl (*MWG Biotech*), *dNTPs* (ATP, CTP, TTP e GTP) 1 mM (*Promega*) e *Taq DNA polimerase FirePol®* a 5 U/μl (*Solis BioDyne*). As condições de amplificação foram: desnaturação inicial a 95 °C, durante 5 minutos, seguida de 30 ciclos de desnaturação a

94 °C, *annealing* a 60 °C, durante 48 segundos e extensão a 72 °C durante 1 minuto seguidos de uma extensão final a 72 °C durante 5 minutos (Pertl *et al.*, 1996).

Tabela 1: Microsatélites autossômicos estudados para a determinação das trissomias (Pertl *et al.*, 1996; Adinolfi *et al.*, 1997).

| <i>Locus</i> | <i>Label</i> | <i>Primers Forward e Reverse</i> | <i>Localização</i> |
|--------------|--------------|--|--------------------|
| D21S11 | FAM | TATGTG AGT CAA TTC CCC AAG TGA GTT GTA TTA GTC AAT GTT CTC CAG | 21q21 |
| D21S1414 | FAM | AAA TTA GTG TCT GGC ACC CAG TA CAA TTC CCC AAG TGA ATT GCC TTC | 21q22.3 |
| D21S1412 | FAM | CGG AGG TTG CAG TGA GTT G GGG AAG GCT ATG GAG GAG A | 21q22.2 |
| D13S631 | HEX | GGC AAC AAG AGC AAA ACT CT TAG CCC TCA CCA TGA TTG G | 13q31-32 |
| D13S634 | HEX | TCC AGA TAG GCA GAT TCA AT CCT TCT TCT TCC CAT TGA TA | 13q14.3 |
| MBPB | HEX | GGA CCT CGT GAA TTA CAA TC ATT TAC CTA CCT GTT CAT CC | 18q23-qter |
| D18S535 | FAM | TCA TGT GAC AAA AGC CAC AC AGA CAG AAA TAT AGA TGA GAA TGC A | 18q12.2 |
| AMXY | FAM | CTG ATG GTT GGC CTC AAG CCT ATG AGG AAA CCA GGG TTC CA | Xp22.3/ Yp11.2 |
| XHPRT | FAM | ATG CCA CAG ATA ATA CAC ATC CCC CTC TCC AGA ATA GTT AGA TGT AGG TAT | Xq26.1 |

2.4.3. Microsatélites do cromossoma Y

Fez-se a determinação dos haplótipos dos indivíduos utilizando entre 8 e 11 microsatélites do cromossoma Y.

A amplificação dos fragmentos de DNA foi feita por PCR, variando as condições consoante os marcadores a estudar. Os *primers* utilizados estão indicados na tabela 2 para 10 microsatélites localizados em várias zonas do cromossoma Y, utilizando-se *primers* marcados com fluorescência tendo sido utilizado quatro *kits* comerciais

nomeadamente o PowerplexY (*Promega*), o Yfiler (*Applied Biosystems*) e o Yplex 5 e Yplex 6 (*Reliagene*).

Tabela 2: Referências dos microsatélites do cromossoma Y utilizados.

| <i>Locus</i> | <i>Label</i> | <i>Primers Forward e Reverse</i> | <i>Referências</i> |
|---------------|--------------|---|--|
| DYS19 | FAM | CTA CTG AGT TTC TGT TAT AGT GGG TTA AGG GA GTG TCA CTA | Kayser <i>et al.</i> (1997) Beleza <i>et al.</i> (2003) |
| DYS385 | HEX | AGC ATG GGT GAC AGA GCT A CCA ATT ACA TAG TCC TCC TTT C | Schneider <i>et al</i> (1998) |
| DYS389 I e II | FAM | CCA ACT CTC ATC TGT ATT ATC TAT TCT TAT CTC CAC CCA CCA GA | Kayser <i>et al.</i> (1997) |
| DYS390 | FAM | TAT ATT TTA CAC ATT TTT GGG CC TGA CAG TAA AAT GAA CAC ATT GC | Kayser <i>et al.</i> (1997) |
| DYS391 | TET | CTA TTC ATT CAA TCA TAC ACC CA CTG GGA ATA AAA TCT CCC TGG TTG CAA G | Kayser <i>et al.</i> (1997) Gusmão <i>et al.</i> (2000) |
| DYS392 | FAM | AAA AGC CAA GAA GGA AAA CAA A AGA CCC AGT TGA TGC AAT GT | Kayser <i>et al.</i> (1997) |
| DYS393 | FAM | GTG GTC TTC TAC TTG TGT CAA TAC AAC TCA AGT CCA AAA AAT GAG G | Kayser <i>et al.</i> (1997) |
| DYS437 | TET | GAC TAT GGG GGT GAG TGC AT AGA CCC TGT CAT TCA CAG ATG A | Gusmão <i>et al.</i> (2002) |
| DYS438 | HEX | TGG GGA ATA GTT GAA CGG TAA GTG GCA GAC GCC TAT AAT CC | Gusmão <i>et al.</i> (2002) |
| DYS439 | TET | TCC TGA ATG GTA CTT CCT AGG TTT GCC TGG CTT GGA ATT CTT TT | Gusmão <i>et al.</i> (2002) |

A mistura de PCR foi a dos *kits* utilizados excepto para os microsatélites separados (DYS388 e DYS460), marcados com fluorescência, em que se utilizaram tampão 25 mM (*Firepol*), MgCl₂ 25 mM (*FirePol*), nucleótidos (dNTPs) a 2 mM (*Promega*) e os primers a 10 pmol/μl (*MWG Biotech*) e 1U Taq-polimerase a 5U/μl (*FirePol Solis Biodyne*).

2.4.4. Estudo da variabilidade do gene *DAZ* do cromossoma Y

Foram utilizados 11 microsatélites do cromossoma Y (tabela 2) que permitiram construir haplótipos de modo a determinar a idade de coalescência e a variabilidade de polimorfismos no cromossoma Y nomeadamente no gene *DAZ* situado na região AZFc, tendo sido utilizadas as mesmas condições descritas em 2.4.3.

2.4.5. Estudo do número de repetições CAG/GGC no gene receptor dos Androgêneos (AR)

Foram utilizadas as condições e os *primers* específicos para dois microsatélites (CAG e GGC) localizados no gene receptor de androgêneos situado no cromossoma X (tabela 3)

Tabela 3: Referências dos *primers* para a amplificação das repetições CAG e GGC do gene AR (Giovannucci *et al.*, 1997).

| <i>Locus</i> | <i>Label</i> | <i>Primers Forward e Reverse</i> | Tamanho (pb) |
|--------------|--------------|---|------------------|
| CAG | FAM | TCC AGA ATC TGT TCC AGA GCG TGC | 252-294 |
| | | GCT GTG AAG GTT GCT GTT CCT CAT | (alelos 14 a 28) |
| GGC | FAM | ACC CTC AGC CGC CGC TTC CTC ATC | 192-246 |
| | | GGA AAG CGA CTT CAC CGC ACC TGA TGT GTG | (alelos 7 a 21) |

Foi utilizada para a determinação do número de repetições GGC a seguinte mistura de PCR: tampão B 10x (*Solis BioDyne*), $MgCl_2$ 25 mM (*Solis BioDyne*), *primers* 10 pmol/ μ l (*MWG Biotech*), DMSO a 15% (Sigma), 7'-deaza-2'-dGTP 2 μ mol (*GEhealthcare*), dNTPs (1 mM de ATP, CTP, TTP e 0,1 mM de GTP (*Promega*) e *Taq* DNA polimerase *FirePol*® a 5 U/ μ l (*Solis BioDyne*) para um volume final de 25 μ l. Para a amplificação do número de repetições CAG foram utilizadas as seguintes condições: tampão B 10x (*Solis BioDyne*), $MgCl_2$ 25 mM (*Solis BioDyne*), *primers* 10 pmol/ μ l (*MWG Biotech*), dNTPs 1 mM (ATP, CTP,

TTP e GTP) (*Promega*) e *Taq* DNA polimerase *FirePol*® a 5 U/μl (*Solis BioDyne*) para um volume final de 25 μl.

2.4.6. Estudo do número de repetições CGG do gene FMR1

Utilizaram-se para este estudo os *primers* NFMR1F (5'-CTCCGTTTCGGTTTCACTTC-3') e NFMR1R (5'-CACCAGCTCCTCCATCTTCT-3') desenhados pelo programa PRIMER3 (<http://frodo.wi.mit.edu>). A PCR foi efectuada segundo Faradz *et al.* (2000) com algumas modificações: desnaturação inicial 3 minutos a 94 °C seguida de 35 ciclos a 94 °C durante 1 minuto, 52 °C, 1 minuto, 72 °C, 1 minuto com uma extensão final de 3 minutos a 72 °C. A mistura da reacção foi a seguinte: tampão B 10x (*Solis BioDyne*), MgCl₂ 25 mM (*Solis BioDyne*), *primers* 10 pmol/μl (*MWG Biotech*), DMSO a 15% (*Sigma*), 7'-deaza-2'-dGTP 2 μmol (*GEhealthcare*), dNTPs (1 mM de ATP, CTP, TTP e 0,1 mM de GTP (*Promega*) e *Taq* DNA polimerase *FirePol*® a 5 U/μl (*Solis BioDyne*) para um volume final de 25 μl.

2.4.7. Estudo da variabilidade de dois polimorfismos do gene IL4

Utilizaram-se para este estudo os resultados da tipagem dos microsatélites CSF1PO e D5S818 obtidos no estudo das populações da Madeira (Publicação I: Fernandes *et al.*, 2002), Açores (Publicação III: Velosa *et al.*, 2002), Cabo Verde (Publicação V: Fernandes *et al.*, 2003) e Guiné-Bissau (Publicação VII: Gonçalves *et al.*, 2002). Analisou-se em primeiro lugar se os 2 polimorfismos do gene *IL4* localizado no cromossoma 5 estavam em LD fazendo o mesmo em relação aos 2 microsatélites utilizados (D5S818 e CSF1PO) e os polimorfismos analisados (IL4590C/T e IL4RP2). Seleccionaram-se os indivíduos que eram duplos heterozigóticos e esses foram excluídos pois não era possível determinar a qual alelo estava ligado cada um dos polimorfismos a estudar. Determinou-se depois a variabilidade dos microsatélites para cada um dos polimorfismos separados e em conjunto uma vez que os dois se encontravam em LD.

2.5. VISUALIZAÇÃO DO DNA

A separação e visualização do DNA foi feito por dois métodos, um manual e outro automático como a seguir se descreve.

2.5.1. Electroforese em gel de acrilamida

As amostras foram colocadas num gel de acrilamida T9C5 (9% de acrilamida e 7% de bis-acrilamida) e sujeitas a uma corrente eléctrica de 150 V e a sua visualização foi feita por nitrato de prata (Budowle *et al.*, 1991), por comparação do tamanho das bandas utilizando como amostras de referência amostras já tipadas noutros laboratórios.

2.5.2. Electroforese capilar

Os fragmentos amplificados foram separados num ABI Prism 310 Genetic Analyser (*Applied Biosystem*) utilizando um capilar de 47 cm e o polímero POP4. Os dados foram analisados utilizando o software Genescan ver 3.1.2 (*Applied Biosystem*). Foram utilizados os seguintes marcadores de tamanho: ILS600 (*Promega*) para os kits Powerplex e Yplex (*Promega*) e o GeneScan-500 TAMRA (*Applied Biosystem*) para os kits Yplex5 e Yplex6 (*Reliagene*) e para os *primers* marcados com fluorescência.

2.6. ANÁLISE ESTATÍSTICA

Foram determinadas para todas as populações as frequências alélicas, o índice de diversidade génica (D_{st}), o índice de fixação (F_{st}), o *Linkage Disequilibrium* (LD), equilíbrio de Hardy-Weinberg (HW), distância genética entre populações e análise de variância molecular (AMOVA, Excoffier *et al.*, 1992) utilizando o programa ARLEQUIN v3.0 (Excoffier *et al.*, 2005; <http://lgb.unige.ch/arlequin/>).

Utilizou-se o programa GENEPOP (Raymond e Rousset, 1995) para calcular a diversidade haplotípica e a diversidade génica segundo Nei (1987).

A comparação entre o número de repetições CAG e GGC foi feita pelo teste exacto de Fisher. Os riscos relativos e os respectivos intervalos de confiança (95%) foram calculados com base na distribuição normal de indivíduos. Valores de $p < 0,05$ foram considerados estatisticamente significativos.

Frequências alélicas

A frequência alélica é calculada pela relação entre o número de vezes que um alelo é encontrado no total de indivíduos.

As frequências alélicas são um parâmetro importante nos estudos de teor evolutivo, pois a variação genética de uma população é normalmente determinada através das diferenças das respectivas frequências alélicas.

Índice de Diversidade Génica (D_{st})

A heterozigotia é uma medida de variação genética de uma população. No entanto, numa população com cruzamentos não ao acaso usa-se o termo diversidade génica. Neste caso, a heterozigotia não é a frequência observada de heterozigóticos na população mas sim a heterozigotia esperada, assumindo que a população está em equilíbrio de Hardy-Weinberg (Nei, 1973). Este parâmetro foi calculado para cada microsatélite para cada população através do programa ARLEQUIN vs 3.0 (Excoffier *et al.*, 2005).

A diversidade génica, equivalente à heterozigotia, corresponde à probabilidade de dois alelos escolhidos ao acaso, num determinado *locus* não estarem presentes no mesmo indivíduo (Nei, 1987). A partir da comparação dos valores de heterozigotia observada e esperada é possível determinar efeitos de *bottleneck* (Cornuet e Luikart, 1996).

Equilíbrio de Hardy-Weinberg

O equilíbrio de Hardy-Weinberg estipula que numa população infinitamente grande, em que os cruzamentos ocorrem ao acaso e sobre o qual não há actuação de factores

evolutivos (selecção, migração ou mutação) as frequências génicas e genotípicas permanecem constantes ao longo das gerações. Este equilíbrio traduz-se na equação $p^2 + 2pq + q^2 = 1$, em que p e q correspondem às frequências dos alelos de um determinado *locus* (Hardy, 1908).

Índice de fixação (F_{st})

Este parâmetro é utilizado para comparar duas populações e pode ser usado como uma medida para estimar as distâncias genéticas entre essas populações. Foi calculado através do programa ARLEQUIN vs. 3.0 (Excoffier *et al.*, 2005). O índice F_{st} tem um valor máximo de 1, sendo que este valor significa uma distância genética máxima entre as populações em comparação (Excoffier *et al.*, 2005).

O cálculo da distância genética e os valores de F_{st} têm em conta as diferenças entre pares de populações (*pairwise difference*), sendo estes valores, uma boa medida do grau de diferenciação genética relativa, dependente não só das frequências haplotípicas como também do valor de Heterozigotia (Reynolds *et al.*, 1983). A distribuição nula refere-se à hipótese de não haver diferenças entre as populações e é obtida permutando os haplótipos entre populações, estabelecendo-se um valor fixo de F_{st} . A probabilidade correspondente refere-se à proporção de permutações nas quais o valor de F_{st} é superior ou igual ao observado. Varia entre 0 e 1, em que 1 se refere a populações que não partilham nenhum dos haplótipos. Valores de R_{st} são análogos aos F_{st} mas baseados no modelo SMM (stepwise Mutation Model; Slatkin, 1995), sendo considerados mais apropriados para estudos com os microsatélites. Valores de R_{st} altos indicam uma grande distância genética entre populações e provavelmente regiões do genoma sujeitos a selecção (Kayser *et al.*, 2003). Estudos de associação entre STRs e outras mutações permitem detectar mutações favoráveis. Se a frequência do alelo seleccionado aumenta, o mesmo é esperado para a frequência do alelo do STR que está no mesmo haplótipo que o alelo sujeito a selecção dependendo, no entanto, do número de alelos e consequentemente da heterozigotia nesse *locus* (Kayser *et al.*, 2003).

Teste exacto de diferenciação populacional

Este teste compara as frequências alélicas absolutas das diferentes populações de modo a determinar a hipótese de distribuição aleatória dos indivíduos nessas populações. Este teste é análogo ao teste exacto de Fisher (tabela de contingência 2x2) (Excoffier *et al.*, 2005) e é definido pela construção de uma tabela em que cada linha corresponde a uma população e cada coluna ao número de haplótipos (Rousset e Raymond, 1995). O teste é considerado significativo se o valor de p for inferior ao nível de significância e nesse caso as duas populações são consideradas significativamente diferentes, utilizando-se quando a amostra é pequena (Excoffier *et al.*, 2005).

Linkage Disequilibrium (LD)

Consiste numa associação não casual de dois ou mais alelos de diferentes *loci* no mesmo cromossoma (Weir, 1996). O *Linkage Disequilibrium* é de primordial importância nos estudos genéticos populacionais pelas implicações biológicas e médicas que possui, sendo de particular importância no mapeamento de genes responsáveis por doenças em humanos (Li e Stephens, 2003). O Delta (D) é a medida absoluta do *Linkage Disequilibrium* e é definido pela diferença entre a frequência haplotípica observada e a frequência haplotípica esperada, tendo em conta a prevalência dos alelos que compõem o haplótipo (Weir, 1996).

Utilizando associações de LD poder-se-á obter um conhecimento da variação do genoma apenas conhecendo uma pequena quantidade de *loci* polimórficos (Goldstein, 2001).

Padrões de LD em genes e em regiões não codificantes podem diferir entre as várias regiões do genoma e dentro das populações (Goldstein, 2001).

Mesmo que os pontos críticos de recombinação sejam comuns no genoma humano, o padrão de LD poderá não ser o mesmo para todas as populações daí a importância de realizar estudos de LD em várias populações (Stumpf e Goldstein, 2003).

Datação de mutações

A idade de coalescência de várias mutações foi estimada a partir da variação de microsatélites do cromossoma Y determinando-se o quadrado da média das diferenças no número de repetições entre todos os cromossomas e o haplótipo fundador (sendo este constituído pelo conjunto dos alelos mais frequentes de cada microsatélite), dividido pela taxa de mutação $W=6,9 \times 10^{-4}$ por 25 anos (Zhivotovsky *et al.*, 2004).

A idade de coalescência foi expressa em milhares de anos (Kya) e a amostra mínima considerada para o cálculo da idade de coalescência foi de seis indivíduos.

Construção de Networks

Utilizou-se o programa NETWORK (Bandelt *et al.*, 1995; 1999) para a determinação de ligações entre os vários haplótipos. Estas ligações ajudam a ilustrar as possíveis relações entre os vários indivíduos com base na variabilidade dos microsatélites do cromossoma Y.

A construção de *Networks* para o cromossoma Y pode ser uma ferramenta útil no estudo da evolução das linhagens paternas. A grande diversidade dos haplótipos indica que este cromossoma não foi alvo de selecção recente apesar das repetições tetranucleotídicas serem muito variáveis e sujeitas a uma alta taxa de mutação (Cooper *et al.*, 1996).

Odds Ratio

O *Odds ratio* (OR) é uma medida utilizada em epidemiologia para avaliar o risco de ocorrência de um resultado particular em indivíduos expostos a um dado factor, relativamente a indivíduos que não foram expostos a esse factor. Pode ser descrito matematicamente como: $OR=AD/BC$, em que A são os casos expostos, B os casos não expostos, C os controlos expostos e D os controlos não expostos (Schlesselman, 1982).

A regressão logística foi utilizada para calcular os valores de OR, os intervalos de confiança (IC) a 95% e os respectivos valores de P. Para valores de OR superiores a 1 considerou-se que nos casos havia uma maior proporção de expostos relativamente aos não expostos. Nos resultados dos estudos de associação, para os Intervalos de Confiança foram consideradas duas possibilidades: não existem diferenças significativas entre casos

e controlos, para IC englobando a unidade e existem diferenças significativas entre casos e controlos para IC não englobando a unidade (Quinn e Keough, 2002).

Determinação de Paternidade

Foram feitos os cálculos para o Índice de paternidade e a probabilidade de paternidade considerando a mãe, um filho e o pressuposto pai (tabela 4) e considerando apenas o filho e o pressuposto pai (tabela 5) (Yepes, 2006), utilizando quatro bases de dados distintas, nomeadamente a da Madeira (Publicação I: Fernandes *et al.*, 2002), Açores (Publicação III: Velosa *et al.*, 2002), Cabo verde (Publicação V: Fernandes *et al.*, 2003) e Guiné (Publicação VII: Gonçalves *et al.*, 2002).

Tabela 4: Fórmulas que permitem o cálculo do índice de paternidade para todos os casos possíveis de genótipos considerando o trio mãe, filho e pressuposto pai.

| Genótipo do pressuposto pai | Genótipo da mãe | Genótipo do filho | X | Y | Índice de Paternidade |
|-----------------------------|-----------------|-------------------|---------------|--------------------|-----------------------|
| AA | AA | AA | 1 | a | 1/a |
| | BB | AB | 1 | a | 1/a |
| | AB | AA | $\frac{1}{2}$ | $\frac{1}{2}a$ | 1/a |
| | AB | AB | $\frac{1}{2}$ | $\frac{1}{2}(a+b)$ | 1/a+b |
| | BC | AB | $\frac{1}{2}$ | $\frac{1}{2}a$ | 1/a |
| AB | AA | AA | $\frac{1}{2}$ | a | 1/2a |
| | AA | AB | $\frac{1}{2}$ | b | 1/2b |
| | CC | AC | $\frac{1}{2}$ | a | 1/2a |
| | AB | AA | $\frac{1}{4}$ | $\frac{1}{2}a$ | 1/2a |
| | AB | AB | $\frac{1}{2}$ | $\frac{1}{2}(a+b)$ | 1/a+b |
| | AC | AA | $\frac{1}{4}$ | $\frac{1}{2}a$ | 1/2a |
| | AC | AC | $\frac{1}{4}$ | $\frac{1}{2}(a+c)$ | $\frac{1}{2}(a+c)$ |
| | AC | AB | $\frac{1}{4}$ | $\frac{1}{2}b$ | 1/2b |
| | AC | BC | $\frac{1}{4}$ | $\frac{1}{2}b$ | 1/2b |
| | CD | AC | $\frac{1}{4}$ | $\frac{1}{2}a$ | 1/2a |

Tabela 5: Fórmulas que permitem o cálculo do índice de paternidade para todos os casos possíveis de genótipos considerando apenas um filho e o pressuposto pai.

| Genótipo do Pressuposto pai | Genótipo do filho | X | Y | Índice de Paternidade |
|--------------------------------|----------------------|--------------------|-------|--------------------------|
| AA | AA | a | a^2 | $1/a$ |
| AA | AB | $1/2a$ | a^2 | $1/2a$ |
| AB | AA | b | $2ab$ | $1/2a$ |
| AB | AB | $\frac{1}{2}(a+b)$ | $2ab$ | $a+b/4ab$ |
| AB | AC | $1/2b$ | $2ab$ | $1/4a$ |

3. RESULTADOS

3.1 ESTUDOS POPULACIONAIS

Foram feitos os estudos das populações do Arquipélago da Madeira, Açores e Cabo Verde utilizando microsátélites autossómicos e microsátélites do cromossoma Y. Posteriormente foram feitas comparações entre as várias populações, utilizando como populações de referência Portugal Continental e Guiné-Bissau.

3.1.1 População da Madeira

Foi determinado o perfil dos indivíduos para vários microsátélites. Alguns indivíduos foram tipados manualmente apenas para cinco microsátélites não havendo por isso informação acerca dos outros microsátélites. Foram calculadas as frequências alélicas de 17 microsátélites para a população da Madeira. Deste estudo concluiu-se que a Madeira não difere estatisticamente do Norte de Portugal apesar de aparecerem, para alguns *loci*, alelos característicos de populações sub-saharianas e norte europeias que não aparecem em Portugal continental (Publicação I: Fernandes *et al.*, 2002).

Posteriormente dividiu-se a Madeira em 4 grupos: Nordeste, Sudoeste, Funchal e Porto Santo, o que permitiu fazer comparações entre estas quatro regiões, não se tendo encontrado diferenças significativas entre elas apesar de se terem encontrado diferenças significativas entre Funchal e Nordeste quando se consideraram quer o estudo do mtDNA quer os SNPs do cromossoma Y (Fernandes *et al.*, 2005). O facto de não se encontrarem diferenças ao nível dos microsátélites autossómicos permitiu juntar todos os indivíduos e definir a população do Arquipélago da Madeira.

Foi também determinado o perfil dos indivíduos para vários microsátélites do cromossoma Y e foram definidos haplótipos e calculadas as frequências alélicas de 7

desses microsatélites para a população da Madeira tendo sido encontrados haplótipos que não se observaram na Península Ibérica mas apenas no Norte da Europa (Publicação II: Fernandes *et al.*, 2001).

3.1.2 População dos Açores

Foi determinado o perfil dos indivíduos para 17 microsatélites apesar de alguns indivíduos não se encontram tipados para todos os marcadores, visto terem falhado as amplificações de alguns marcadores e em vez de repeti-los optou-se por aumentar o tamanho da amostra introduzindo novos indivíduos. Foram calculadas as frequências alélicas para a população dos Açores e quando se fez a comparação com a população do Norte de Portugal observou-se que não diferem estatisticamente, excepto com o Norte de Portugal, para o *locus* D18S51, e com o Centro de Portugal para o *locus* VWA. Quando se comparou com a população da Madeira observam-se diferenças significativas para os *loci* D5S818, D13S317, FGA, VWA e CD4, surgindo na população dos Açores alguns alelos que não existem nem na Madeira nem em Portugal Continental (Publicação III: Velosa *et al.*, 2002).

Posteriormente fez-se a divisão dos Açores em 3 grupos sendo o primeiro constituído pelo grupo Oriental que engloba São Miguel e Santa Maria, o segundo pelo grupo Central que engloba a Terceira, São Jorge, Pico Horta e Faial e o terceiro grupo, o grupo Ocidental que é constituído por Graciosa, Flores e Corvo. Realizou-se o teste de *Fst* e como não se observaram diferenças significativas ($p > 0,05$) entre nenhum dos grupos juntaram-se todos os indivíduos num só grupo que se utilizou para a comparação entre as várias populações dos vários arquipélagos considerados.

Foi também determinado o perfil dos indivíduos para 10 microsatélites do cromossoma Y e foram definidos haplótipos e calculadas as frequências alélicas desses microsatélites para a população dos Açores. Foram encontradas diferenças significativas com a Madeira para os *loci* DYS390, DYS391 e DYS392 e com Portugal Continental para os *loci* DYS391 e DYS393. Quando foram analisadas as frequências haplotípicas observaram-se diferenças significativas para a população dos Açores quando comparada com a Madeira e com Portugal Continental (Publicação IV: Fernandes e Brehm, 2003).

3.1.3 População de Cabo Verde

Foi determinado o perfil dos indivíduos para 17 microsátélites apesar de alguns indivíduos não se encontrarem tipados para todos os marcadores, visto terem falhado as amplificações de alguns marcadores e em vez de repeti-los optou-se por aumentar o tamanho da amostra, introduzindo novos indivíduos. Foram calculadas as frequências alélicas para duas populações de Cabo Verde, uma englobando o grupo do Barlavento e outra o Sotavento com base em estudos anteriores de DNA mitocondrial (Brehm *et al.*, 2002), onde se que demonstrou haver diferenças significativas entre os dois grupos.

Fez-se uma primeira análise de *Fst* considerando todas as ilhas separadamente mas atendendo ao facto de não haver diferenças significativas entre algumas ilhas foram definidos dois grupos: o primeiro, o grupo do Barlavento, constituído por São Vicente, São Nicolau e Santo Antão e o segundo, o grupo do Sotavento, constituído por Santiago, Fogo, Brava, Maio, Boavista e Sal (Publicação V: Fernandes *et al.*, 2003).

3.2 BASES DE DADOS POPULACIONAIS

Com os dados obtidos foram construídas bases de dados populacionais, quer para os microsátélites autossómicos quer para os microsátélites do cromossoma Y.

3.2.1 Bases de dados dos microsátélites autossómicos

Cinco bases de dados populacionais são apresentadas como referência para estudos de populações: Madeira (Publicação I: Fernandes *et al.*, 2002), Açores (Publicação III: Velosa *et al.*, 2003), Cabo Verde (Publicação V: Fernandes *et al.*, 2003) e Guiné-Bissau (Publicação VII: Gonçalves *et al.*, 2002). A base de dados da população da Madeira é utilizada como referência para a averiguação de paternidades em que a frequência dos alelos de cada um dos marcadores é essencial para os cálculos.

Determinou-se o perfil de cada um dos indivíduos envolvido no caso de paternidade utilizando 13 marcadores autossómicos comuns a todas as bases de dados (tabela 6).

Tabela 6: Microsatélites e perfis utilizados para um caso de paternidade apenas com o Pressuposto pai e um filho.

| Microsatélite | Filho | Pressuposto pai |
|---------------|----------|-----------------|
| CSF1PO | 11, 11 | 11,11 |
| D13S317 | 8, 13 | 8, 12 |
| D16S539 | 9, 10 | 9, 13 |
| D18S51 | 12, 14 | 12, 13 |
| D21S11 | 30, 31.2 | 29, 31.2 |
| D3S1358 | 15, 18 | 15, 16 |
| D5S818 | 12, 13 | 12, 12 |
| D7S820 | 10, 13 | 8, 10 |
| D8S1179 | 13, 15 | 15, 16 |
| FGA | 24, 24 | 24, 24 |
| TH01 | 7, 9.3 | 7, 8 |
| TPOX | 11, 11 | 11, 11 |
| VWA | 18, 18 | 18, 18 |

Fez-se o cálculo do IP (índice de Paternidade) e de W (probabilidade de paternidade), utilizando apenas os 13 microsatélites estudados para todas as populações (tabela 7).

Apesar da importância das bases de dados populacionais, conseguiu-se demonstrar, utilizando apenas 13 marcadores, que no caso testado o resultado em relação à paternidade é o mesmo quer se utilize uma base de dados da Madeira quer se utilize uma completamente diferente como a da Guiné-Bissau (tabela 7).

Tabela 7: Determinação do Índice de Paternidade (IP) e da Probabilidade de Paternidade (W) utilizando, para o mesmo caso 5 bases de dados distintas.

| Microsatélite | PI Madeira | PI Guiné | PI Açores | PI CVB | PI CVS |
|---------------|------------|-----------|------------|------------|-------------|
| CSF1PO | 2,8986 | 5,8824 | 3,1746 | 3,4965 | 4,8780 |
| D13S317 | 2,9412 | 16,6667 | 2,3585 | 3,8462 | 4,3860 |
| D16S539 | 2,0000 | 1,3889 | 1,9231 | 0,9434 | 1,2077 |
| D18S51 | 1,2500 | 3,3333 | 1,9231 | 4,2373 | 3,9063 |
| D21S11 | 3,3333 | 4,5455 | 2,6042 | 4,7170 | 4,9020 |
| D3S1358 | 0,8333 | 0,8475 | 1,0000 | 1,2626 | 2,4510 |
| D5S818 | 1,3158 | 1,2346 | 1,3928 | 1,2165 | 1,1468 |
| D7S820 | 0,9804 | 0,6579 | 0,9804 | 0,5828 | 0,6831 |
| D8S1179 | 1,9231 | 1,1628 | 2,3585 | 2,0492 | 1,5723 |
| FGA | 7,4074 | 9,0909 | 7,7519 | 5,8480 | 5,3476 |
| TH01 | 1,7483 | 0,7143 | 1,3587 | 0,8961 | 1,0204 |
| TPOX | 3,8462 | 3,0769 | 3,3223 | 3,8023 | 3,0395 |
| VWA | 5,7143 | 10,5263 | 5,6497 | 7,0922 | 6,9444 |
| | PI total | PI total | PI total | PI total | PI total |
| | 41800,4728 | 347285,49 | 45907,9520 | 65726,0630 | 172032,2659 |
| | W | W | W | W | W |
| | 99,9976 | 99,9997 | 99,9978 | 99,9985 | 99,9994 |

3.2.2 Bases de dados dos microsatélites do cromossoma Y

Duas bases de dados do cromossoma Y são apresentadas, uma para a população da Madeira (Publicação II: Fernandes *et al.*, 2001) e outra para a população dos Açores (Publicação IV: Fernandes e Brehm, 2003). Na base de dados da Madeira são apresentadas as frequências alélicas de 7 microsatélites do cromossoma Y e também os haplótipos constituídos pelos mesmos. Na base de dados dos Açores são apresentadas frequências alélicas para cada um dos microsatélites estudados e também as frequências haplotípicas utilizando 10 microsatélites. Parâmetros como diversidade génica e diversidade haplotípica são apresentados.

3.3 ASSOCIAÇÃO A DOENÇAS

Vários estudos associados a doenças foram realizados utilizando microsatélites. Utilizando microsatélites autossômicos neutrais para os cromossomas 13, 18 e 21, fez-se a detecção da presença de aneuploidias através do líquido amniótico. Através do estudo de microsatélites, constituídos por repetições trinucleotídicas, nomeadamente repetições CAG e GGC no gene *AR*, determinou-se a variação alélica em casos de indivíduos com níveis de PSA elevados e em indivíduos com infertilidade masculina. Foram ainda analisadas as repetições GGC do gene *FMR1* para o Síndrome de X-frágil.

3.3.1 Detecção de trissomias por PCR

Dos 120 casos analisados apenas em 2 se verificou a existência de trissomia 21. Os resultados destes dois casos foram inequívocos aparecendo três fragmentos com diferentes tamanhos, um para cada cromossoma 21. Na maior parte dos casos não existiam aneuploidias e obtiveram-se dois fragmentos para cada indivíduo. Em alguns casos o aparecimento de apenas um pico não permitiu concluir se existiam ou não aneuploidias pois na maior parte destes casos os indivíduos eram homozigóticos para ambos os marcadores desse cromossoma. Nalguns casos surgiram apenas dois fragmentos apresentando um dos picos o dobro da altura do outro. Nestes casos suspeitou-se de trissomia. No entanto ao analisar os outros marcadores do mesmo cromossoma, o mesmo já não se verificava.

3.3.2 Repetições GGC e CAG do gene *AR* em indivíduos com níveis de PSA elevados

Foi determinado o genótipo para cada indivíduo sendo a designação de cada alelo determinado pelo número de repetições. Foram calculadas as frequências alélicas para as repetições GGC e CAG (tabela 8).

Tabela 8: Frequências alélicas de repetições GGC e CAG do gene AR para duas populações, uma de indivíduos com níveis de PSA elevados (PSAd) e outra com níveis de PSA normais (PSAc).

| GGC | PSAd (n=96) | PSAc (n=94) | CAG | PSAd (n=96) | PSAc (n=94) |
|-----|----------------|----------------|-----|----------------|----------------|
| 3 | | | 10 | | |
| 6 | | | 14 | 0,010 | |
| 8 | 0,010 | | 15 | | 0,011 |
| 10 | | 0,011 | 16 | 0,042 | 0,011 |
| 11 | 0,010 | | 17 | 0,042 | 0,032 |
| 12 | | 0,011 | 18 | 0,021 | 0,043 |
| 13 | 0,031 | 0,043 | 19 | 0,219 | 0,181 |
| 14 | 0,010 | 0,011 | 20 | 0,125 | 0,213 |
| 15 | 0,031 | | 21 | 0,198 | 0,128 |
| 16 | 0,010 | 0,011 | 22 | 0,125 | 0,138 |
| 17 | 0,469 | 0,489 | 23 | 0,052 | 0,074 |
| 18 | 0,385 | 0,351 | 24 | 0,083 | 0,085 |
| 19 | 0,042 | 0,043 | 25 | 0,031 | 0,064 |
| 20 | | 0,021 | 26 | 0,042 | 0,021 |
| 21 | | 0,011 | 27 | 0,010 | |
| 22 | | | 28 | | |
| 26 | | | 29 | | |

Quando calculado o *Odds Ratio* para os vários intervalos não foram encontradas correlações nem quando foram considerados os valores das frequências em separado nem quando se fez a associação das várias combinações GGC/CAG.

3.3.3 Repetições GGC e CAG do gene AR em indivíduos com infertilidade masculina

Foi determinado o genótipo para cada indivíduo sendo a designação de cada alelo determinado pelo número de repetições. Foram calculadas as frequências alélicas para as repetições GGC e CAG (tabela 9).

Tabela 9: Frequências alélicas de repetições GGC e CAG do gene AR para duas populações, uma de homens inférteis e outra de homens férteis.

| | Inférteis (n=108) | Férteis (n=99) | | Inférteis (n=108) | Férteis (n=99) |
|------------|----------------------|-------------------|------------|----------------------|-------------------|
| GGC | | | CAG | | |
| 3 | 0,028 | 0,010 | 10 | | 0,010 |
| 6 | 0,009 | | 17 | 0,009 | 0,020 |
| 10 | | 0,010 | 18 | 0,093 | 0,020 |
| 11 | 0,019 | | 19 | 0,083 | 0,070 |
| 12 | 0,009 | 0,010 | 20 | 0,194 | 0,120 |
| 13 | 0,046 | 0,020 | 21 | 0,148 | 0,130 |
| 14 | 0,019 | 0,010 | 22 | 0,120 | 0,150 |
| 15 | 0,028 | 0,030 | 23 | 0,111 | 0,110 |
| 16 | 0,019 | 0,010 | 24 | 0,093 | 0,130 |
| 17 | 0,491 | 0,550 | 25 | 0,056 | 0,090 |
| 18 | 0,250 | 0,260 | 26 | 0,056 | 0,060 |
| 19 | 0,056 | 0,050 | 27 | 0,019 | 0,030 |
| 20 | 0,019 | 0,020 | 28 | 0,009 | 0,050 |
| 21 | | 0,010 | 29 | 0,009 | 0,010 |
| 22 | 0,009 | | | | |
| 26 | | 0,010 | | | |

Quando foi calculado o *Odds Ratio* (OD) para os vários alelos das repetições CAG e GGC, não foram encontradas diferenças significativas. No entanto, quando se consideraram intervalos de variação do tamanho dos alelos, foram encontradas algumas correlações, nomeadamente para os alelos CAG<21 obteve-se um valor de OR = 1,9123 (CI 95%, 1,0475-3,4912), indicando que estes alelos poderão conferir susceptibilidade para a infertilidade, não sendo evidente nenhuma associação com as repetições GGC.

No entanto, quando se fez a associação das várias combinações GGC/CAG, obteve-se, para a combinação GGC>16/CAG>23, um valor de OR=0,5227 (CI 95% 0,2735-0,9989) que evidencia um possível efeito protector para alelos maiores dos dois marcadores, o que não se observou na análise dos marcadores em separado.

3.3.4 Comparação entre o número de repetições CAG e GGC do gene AR

Foi feita a comparação entre o número de repetições CAG e GGC do gene AR para 4 grupos: indivíduos inférteis e indivíduos férteis (tabela 10); indivíduos com níveis de PSA elevados e com níveis de PSA normais (tabela 11).

Quando se compararam as quatro populações de indivíduos pelo teste de *Fst*, encontraram-se diferenças significativas quando comparada a população de indivíduos com níveis de PSA elevado com os inférteis ($0,00901 \pm 0,0091$) bem como destes com os férteis ($0,01802 \pm 0,0182$). Atendendo a que não há diferenças significativas entre as populações de férteis e inférteis, originárias do Norte de Portugal, e entre os indivíduos com níveis de PSA elevado e os que apresentam níveis normais, originários da Madeira, comparou-se então os dois grupos populacionais de Norte de Portugal ($n=207$) e Madeira ($n=190$) tendo-se encontrado diferenças significativas entre as duas populações que não se deverão à associação com nenhuma patologia mas que indicam uma diferença entre as duas populações já observada quer para estudos do DNA mitocondrial (Brehm *et al.*, 2003), quer em estudos noutros *loci* autossómicos como os HLA's (Spínola *et al.*, 2005).

Tabela 10: Frequências de vários intervalos de comparação CAG e GGC para indivíduos inférteis e indivíduos férteis

| Inférteis CAG | GGC ≤16 | 17 | 18 | ≥19 | total |
|------------------|--------------|--------------|--------------|--------------|--------------|
| ≤20 | 0,046 | 0,148 | 0,139 | 0,046 | 0,380 |
| 21 | 0,019 | 0,074 | 0,046 | 0,009 | 0,148 |
| 22 | 0,028 | 0,074 | 0,019 | | 0,120 |
| 23 | 0,028 | 0,037 | 0,037 | 0,009 | 0,111 |
| ≥24 | 0,056 | 0,157 | 0,009 | 0,019 | 0,241 |
| total | 0,176 | 0,491 | 0,250 | 0,083 | |

| Férteis CAG | GGC ≤16 | 17 | 18 | ≥19 | total |
|----------------|--------------|--------------|--------------|--------------|--------------|
| ≤20 | 0,020 | 0,081 | 0,081 | 0,061 | 0,242 |
| 21 | 0,010 | 0,061 | 0,040 | 0,020 | 0,131 |
| 22 | | 0,111 | 0,040 | | 0,152 |
| 23 | | 0,081 | 0,020 | 0,010 | 0,111 |
| ≥24 | 0,061 | 0,222 | 0,071 | 0,010 | 0,364 |
| total | 0,091 | 0,556 | 0,253 | 0,101 | |

Tabela 11: Frequências de vários intervalos de comparação CAG e GGC para indivíduos: com níveis de PSA elevados (PSAd) e com níveis de PSA normais (PSAc).

| PSAd | GGC | | | | |
|--------------|--------------|--------------|--------------|--------------|--------------|
| CAG | ≤16 | 17 | 18 | ≥19 | total |
| ≤20 | 0,063 | 0,177 | 0,188 | 0,031 | 0,458 |
| 21 | 0,010 | 0,115 | 0,073 | | 0,198 |
| 22 | 0,021 | 0,073 | 0,031 | | 0,125 |
| 23 | | 0,010 | 0,042 | | 0,052 |
| ≥24 | 0,010 | 0,094 | 0,052 | 0,010 | 0,167 |
| total | 0,104 | 0,469 | 0,385 | 0,042 | |

| PSAc | GGC | | | | |
|--------------|--------------|--------------|--------------|--------------|--------------|
| CAG | ≤16 | 17 | 18 | ≥19 | total |
| ≤20 | 0,064 | 0,181 | 0,191 | 0,053 | 0,489 |
| 21 | | 0,085 | 0,043 | | 0,128 |
| 22 | 0,011 | 0,074 | 0,053 | | 0,138 |
| 23 | | 0,032 | 0,032 | 0,011 | 0,074 |
| ≥24 | 0,011 | 0,117 | 0,032 | 0,011 | 0,170 |
| total | 0,085 | 0,489 | 0,351 | 0,074 | |

3.3.5 Repetições GGC no gene FMR1 para a população da Madeira e Guiné-Bissau

Foi determinado o genótipo para cada indivíduo das populações da Madeira e da Guiné-Bissau, sendo a designação de cada alelo determinado pelo número de repetições (tabela 12).

Fez-se a análise dos resultados calculando a frequência de cada um dos alelos para as duas populações em estudo e quando se compararam as duas populações observou-se a existência de diferenças significativas ($P < 0,05$).

Tabela 12: Distribuição alélica de repetições GGC do gene FMR1 para as populações da Madeira (n=141) e da Guiné-Bissau (n=104).

| (CGG)n | Madeira (n=141) | Guiné-Bissau (n=104) |
|-----------|--------------------|-------------------------|
| 7 | 0,007 | - |
| 8 | 0,007 | 0,009 |
| 13 | 0,007 | - |
| 14 | 0,035 | - |
| 16 | 0,007 | - |
| 19 | 0,028 | - |
| 20 | 0,071 | 0,009 |
| 21 | 0,028 | 0,009 |
| 22 | - | 0,019 |
| 23 | 0,064 | 0,077 |
| 24 | 0,014 | 0,019 |
| 25 | 0,014 | 0,019 |
| 27 | 0,007 | 0,009 |
| 28 | 0,100 | 0,067 |
| 29 | 0,262 | 0,231 |
| 30 | 0,248 | 0,221 |
| 31 | 0,057 | 0,144 |
| 32 | 0,021 | 0,019 |
| 33 | - | 0,009 |
| 34 | - | 0,019 |
| 35 | - | 0,019 |
| 36 | 0,007 | 0,009 |
| 38 | 0,014 | 0,038 |
| 39 | - | 0,009 |
| 40 | - | 0,009 |
| 41 | - | 0,029 |
| 43 | | 0,009 |
| He | 0,849 | 0,870 |

Determinou-se posteriormente a frequência desses alelos em duas subpopulações da população da Madeira, uma de indivíduos sobredotados e outra de indivíduos com suspeita de Síndrome de X-frágil (tabela 13).

Tabela 13: Distribuição alélica de repetições GGC do gene FMR1 para três grupos da população da Madeira: população em geral (n=115), grupo de sobredotados (n=26) e indivíduos com suspeita de X-frágil (n=25).

| (CGG)n | Madeira (n=115) | Sobredotados (n=26) | X-frágil (n=25) |
|--------|---------------------------|---------------------------|---------------------------|
| 7 | - | 0,038 | - |
| 8 | 0,009 | - | - |
| 13 | 0,009 | - | - |
| 14 | 0,034 | 0,038 | - |
| 16 | 0,009 | - | - |
| 19 | 0,034 | - | - |
| 20 | 0,061 | 0,116 | - |
| 21 | 0,026 | 0,038 | - |
| 23 | 0,052 | 0,116 | - |
| 24 | 0,009 | 0,038 | - |
| 25 | 0,017 | - | - |
| 27 | 0,009 | - | 0,080 |
| 28 | 0,122 | - | 0,080 |
| 29 | 0,270 | 0,231 | 0,520 |
| 30 | 0,235 | 0,308 | 0,120 |
| 31 | 0,052 | 0,077 | 0,080 |
| 32 | 0,026 | - | 0,040 |
| 33 | - | - | - |
| 34 | - | - | 0,040 |
| 35 | - | - | - |
| 36 | 0,009 | - | - |
| 38 | 0,017 | - | 0,040 |
| He | 0,8508 (0,0195) | 0,8462 (0,0448) | 0,7200 (0,0904) |

Observaram-se diferenças significativas entre a população da Madeira e a população com suspeita de Síndrome de X-frágil, não se tendo observado diferenças desta com a população de sobredotados, havendo diferenças significativas entre a população de sobredotados e a população de indivíduos com suspeita de Síndrome de X-frágil.

Estas diferenças devem-se à elevada frequência do alelo 29 nos indivíduos com suspeita de Síndrome de X-frágil (figura 4), indicando uma possível ligação deste alelo a

outras mutações no cromossoma X que poderão ser responsáveis pelo fenótipo de Síndrome de X-frágil nestes indivíduos.

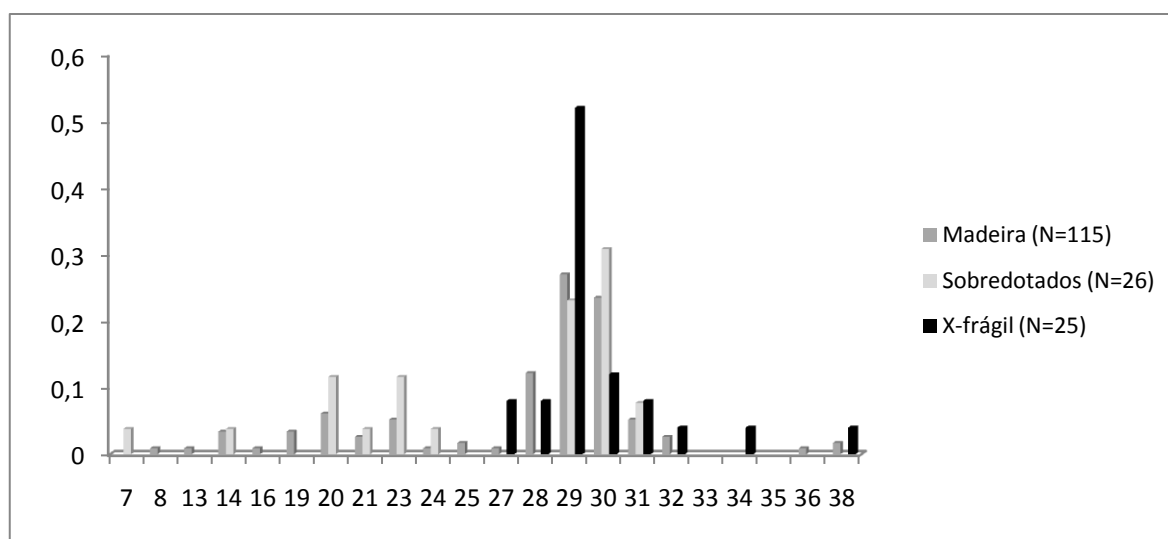


Figura 4: Distribuição dos alelos GGC nas três subpopulações da Madeira

3.4 ANÁLISE DE MUTAÇÕES

O estudo dos microsatélites do cromossoma Y associados a SNPs (*Single Nucleotide Polymorphisms*), determinados por Gonçalves e colaboradores (Gonçalves *et al.*, 2003; 2005), permitiu estabelecer a variabilidade dos microsatélites e a idade de coalescência de todos os haplogrupos do cromossoma Y, nas populações da Madeira e dos Açores. Na Madeira e nos Açores dois possíveis efeitos fundadores são identificados nomeadamente em relação aos haplogrupos J2 e G respectivamente (Publicação IX: Fernandes *et al.*, 2008b).

Determinou-se a variabilidade de microsatélites do cromossoma Y para o haplogrupo E3b de indivíduos das populações da Madeira, Açores, Portugal Continental, Guiné-Bissau e Cabo Verde sendo possível definir um haplótipo associado a cada um dos sub-haplogrupos dentro deste haplogrupo, detectando-se também possíveis mutações em alguns SNPs do cromossoma Y. Foi ainda possível fazer datações de cada um dos sub-haplogrupos do E3b e comparar com outros estudos realizados anteriormente para outras populações (Publicação VIII: Fernandes *et al.*, 2008a).

Fez-se a datação de mutações no cromossoma Y, nomeadamente no gene *DAZ*, relacionado com a infertilidade (Publicação IX: Fernandes *et al.*, 2006). Através da diversidade observada em dois microsatélites autossómicos do cromossoma 5 foi possível determinar quais as combinações mais antigas constituídas pelos polimorfismos RP2 e 590 C/T do gene *IL4*. Foi ainda possível determinar fenómenos de crescimento rápido de algumas populações bem como a possível ocorrência de selecção nesse gene, em algumas populações (Publicação XI: Berenguer *et al.*, *submitted*).

3.4.1 Cromossoma Y

No estudo dos haplogrupos do cromossoma Y e do DNA mitocondrial para a população da Madeira verificou-se que havia diferenças significativas ($P < 0,05$) entre os grupos de Nordeste e Sudoeste e Nordeste e Funchal bem como o Funchal com o Porto Santo (Fernandes *et al.*, 2005) mas, como não se verificaram diferenças significativas entre Nordeste e Porto Santo e entre Sudoeste e Funchal, juntaram-se os indivíduos em apenas dois grupos que se designaram por Nordeste e Sudoeste. Fez-se o estudo dos microsatélites do cromossoma Y e apesar de existirem diferenças significativas quer para os haplogrupos do cromossoma Y quer para o mtDNA, o mesmo não se verificou nem para os microsatélites autossómicos nem para os do cromossoma Y. Verificou-se, no entanto, a existência de um efeito fundador em relação à linhagem J do cromossoma Y, no Porto Santo (Publicação IX: Fernandes *et al.*, 2008b).

3.4.2 Gene *DAZ*

O gene *DAZ* está relacionado com a infertilidade masculina. Os homens podem ter no seu genoma até 4 cópias deste gene, situado no cromossoma Y, observando-se uma variação do número de cópias de indivíduo para indivíduo, o que torna possível a definição de haplótipos para este gene. Alguns homens inférteis apresentam microdelecções em algumas cópias deste gene. No entanto foi estabelecida uma relação entre essas microdelecções e os haplogrupos do cromossoma Y, tendo-se concluído que

muitas das variações observadas constituem polimorfismos na população fértil. Recorrendo à variabilidade dos STRs do cromossoma Y foi possível determinar, segundo Zhivotovsky *et al.* (2004), a idade de coalescência dos vários haplótipos do gene *DAZ* (Publicação X: Fernandes *et al.*, 2006).

3.4.3 Gene *IL4*

Utilizaram-se para este estudo os resultados da tipagem dos microsatélites CSF1PO e D5S818 obtidos no estudo de populações. Analisou-se em primeiro lugar se os 2 polimorfismos do gene *IL4* localizado no cromossoma 5 estavam em LD, fazendo o mesmo em relação aos 2 microsatélites utilizados e os polimorfismos analisados. Depois de se verificar a existência de LD, determinou-se a variabilidade dos microsatélites para cada um dos polimorfismos. Fez-se o mesmo depois de se definirem haplótipos utilizando os dois polimorfismos (Publicação XI: Berenguer *et al.*, *submitted*).

4. DISCUSSÃO E CONCLUSÕES

Os microsatélites permitem-nos estudar o genoma quer numa perspectiva do indivíduo quer em relação a diferentes populações.

Nos vários estudos cuja compilação resulta nesta tese, caracterizaram-se geneticamente várias populações humanas dos Arquipélagos da Madeira, Açores e Cabo Verde tendo sido construídas bases de dados de microsatélites autossómicos para cada uma das populações estudadas. Verificou-se que existem diferenças significativas entre as várias ilhas de Cabo Verde (Publicação V: Fernandes *et al.*, 2003).

Foram definidas idades de coalescência a partir do estudo dos microsatélites do cromossoma Y, que permitiram concluir que as cópias do gene *DAZ*, situado no cromossoma Y, são o resultado de um processo evolutivo estando essa evolução associada a alguns haplogrupos. Verificou-se também a ocorrência de possíveis mutações nos SNPs que definem os haplogrupos através da comparação dos microsatélites do cromossoma Y dentro de cada haplogrupo especialmente no haplogrupo E3b (Publicação VIII: Fernandes *et al.*, 2008a).

Verificou-se existir uma associação entre o número de repetições CAG e GGC do gene Receptor de Androgéneos (AR), situado no cromossoma X e a infertilidade, especialmente quando combinados os dois polimorfismos, parecendo haver um efeito protector dos alelos maiores e alguma susceptibilidade para os alelos menores.

Quando se estudou o número de repetições GGC do gene *FMR1* em doentes com suspeita de Síndrome de X-frágil observaram-se diferenças significativas quando comparadas com a população em geral e com um grupo de sobredotados. Essa diferença deveu-se principalmente à presença do alelo 29 em quase todos os indivíduos o que, por si só, não constitui um factor de risco mas poderá ser uma indicação da associação deste alelo com outra mutação no mesmo gene, responsável por este fenótipo (Kunst *et al.*, 1996).

Os microsatélites podem tornar-se numa preciosa ferramenta no estudo de outras mutações no mesmo cromossoma, principalmente se se tratar de mutações no cromossoma Y.

Utilizaram-se os microsatélites do cromossoma Y para definir haplótipos, associá-los aos haplogrupos (definidos por mutações pontuais) e estudar a influência das linhagens paternas nos três arquipélagos.

É de salientar a importância dos microsatélites na datação das mutações. Com esta metodologia foi possível provar que os haplótipos do gene *DAZ* resultam da evolução e não parecem estar relacionados com a infertilidade masculina (Publicação IX: Fernandes *et al.*, 2006).

4.1 ESTUDOS POPULACIONAIS

Através do estudo das populações foi possível verificar que, apesar da colonização dos três arquipélagos ter sido feito pelos portugueses, existem diferenças significativas entre os vários arquipélagos e dentro destes existem diferenças que reflectem as várias influências durante e após a colonização.

Recorrendo aos estudos dos microsatélites autossómicos foi possível identificar diferenças entre os diferentes grupos de ilhas de Cabo Verde (Publicação V: Fernandes *et al.*, 2003), que foram posteriormente confirmadas através da análise do cromossoma Y (Gonçalves *et al.*, 2003) e do mtDNA (Brehm *et al.*, 2002). Identificaram-se várias influências na população de Cabo Verde, nomeadamente uma influência europeia por linhagens paternas (observada através do cromossoma Y) de cerca de 50% (Gonçalves *et al.*, 2003) que não se verifica por linhagens maternais (DNA mitocondrial) (Brehm *et al.*, 2002).

Apesar de se terem encontrado evidências de uma micro-diferenciação da população da Madeira através do estudo do cromossoma Y e do mtDNA, que poderá ser o reflexo do processo de divisão e povoamento inicial do arquipélago (Fernandes *et al.*, 2005), o mesmo não se observou quando se recorreu ao estudo dos microsatélites autossómicos.

O estudo do cromossoma Y nas populações da Madeira e dos Açores permitiu observar diferenças nas influências masculinas nos dois arquipélagos, recorrendo à variabilidade de

microsatélites do cromossoma Y dentro dos haplogrupos, detectando-se dois possíveis efeitos de fundador (Publicação VIII: Fernandes *et al.*, 2008a).

A elevada taxa de mutação dos microsatélites faz com que estes sejam bons marcadores genéticos quando se pretende a diferenciação entre indivíduos e, quando se tratam de microsatélites neutrais, podem ser um auxiliar precioso na identificação de indivíduos e no estudo de populações.

Utilizou-se o estudo de microsatélites autossómicos neutrais do Sul de Portugal para justificar o aparecimento de um desequilíbrio de *Hardy-Weinberg* nesta população para a mutação CCR5 $\Delta 32$, concluindo-se que esta esteve sujeita a selecção, visto que este desequilíbrio não se verifica no estudo dos microsatélites (Freitas *et al.*, 2006).

4.2 BASES DE DADOS E TESTES DE PATERNIDADE

O estudo dos microsatélites autossómicos permitiu a criação de bases de dados das três populações. Apesar das bases de dados populacionais serem importantes, demonstrou-se que não é imprescindível a utilização de uma base de dados específica da população, para os cálculos de paternidade, podendo ser utilizados quaisquer valores de frequência dos vários microsatélites. A partir de apenas 13 microsatélites, não havendo nenhuma exclusão, o valor de cada um dos IP (Índice de paternidade) será irrelevante para a conclusão final da paternidade, desde que seja este valor seja superior a 10000, o que corresponde a uma probabilidade de paternidade superior a 99,99%.

As bases de dados populacionais assumem principalmente uma importância primordial na identificação de indivíduos, a partir de qualquer amostra biológica, e em estudos populacionais onde se pretende comparar quer perfis de indivíduos quer frequências alélicas das várias populações em estudo.

4.3 ASSOCIAÇÃO A DOENÇAS

O método de detecção rápida de aneuploidias a partir de líquido amniótico, mostrou ser rápido e eficaz na de detecção de aneuploidias. Para melhorar os resultados podem

ser utilizados outros microsatélites dos cromossomas 13, 18 e 21 que permitiriam esclarecer qualquer dúvida. Esta técnica pode vir a ser ainda mais útil, quando for possível fazer diagnóstico pré-natal utilizando métodos não invasivos, a partir do sangue materno (Samura *et al.*, 2001). Também permite detectar se a origem do alelo supranumerário é materno ou paterno, por comparação dos alelos dos progenitores.

O estudo do número de repetições GGC no gene *FMR1* (*Fragile X Mental Retardation*), situado no cromossoma X e associado à Síndrome de X-frágil, permitiu inferir acerca da probabilidade da ocorrência de indivíduos pré-mutados e doentes na população da Madeira. A observação da existência de uma baixa frequência de alelos de tamanho elevado, pode justificar a inexistência na Madeira de nenhum indivíduo com diagnóstico confirmado de Síndrome de X-frágil, devido a mutação no gene *FMR1*. No entanto, a predominância do alelo 29 nos indivíduos com diagnóstico clínico de Síndrome de X-frágil, quando comparado com a população em geral ($P < 0,05$), poderá indicar uma associação a outra mutação neste gene ou noutro gene, como o *FRAXE*, que se encontre em LD e que poderá ser responsável pelo fenótipo observado nestes indivíduos.

A partir da determinação da frequência dos alelos dos tripletos CAG e GGC do gene *AR* (*Androgen Receptor*) do cromossoma X, foi possível comparar os valores obtidos para um grupo de indivíduos com infertilidade masculina e os de outro constituído por indivíduos que apresentam níveis de PSA elevados. Definindo haplótipos pode-se fazer também a comparação entre os mesmos grupos. Apesar de não haver diferenças significativas entre os dois grupos de férteis e inférteis quando se consideram intervalos de repetições já se observam diferenças significativas indicando que fragmentos com tamanhos pequenos quer de repetições CAG quer GGC poderão predispor para a infertilidade enquanto os tamanhos maiores, principalmente de repetições CAG, poderão ter um efeito protector em relação à infertilidade. Considerando que os indivíduos do estudo férteis/inférteis eram oriundos de uma população do Norte de Portugal e os indivíduos do estudo dos níveis de PSA eram da Madeira, quando se comparam os dois grupos populacionais observam-se diferenças significativas entre eles o que vem reforçar a importância de manter, para estudos de caso controlo, populações da mesma região e não apenas do mesmo país onde pode haver subestruturação da população, como parece ter acontecido neste caso.

4.4 ANÁLISE DE MUTAÇÕES

A variabilidade dentro dos microsátélites e a construção de diagramas em rede (*networks*) permitiram detectar fenómenos de efeito fundador.

As datações do haplogrupo E3b permitiram verificar efeitos fundadores para este haplogrupo, no caso de Cabo Verde e Portugal Continental (Publicação VII: Fernandes *et al.*, 2008a), bem como nos casos dos haplogrupos J2 e G na Madeira e Açores, respectivamente (Publicação IX: Fernandes *et al.*, 2008b).

Através da combinação dos resultados dos microsátélites (haplótipos) e SNPs (haplogrupos) do cromossoma Y observou-se a existência de mutações recorrentes nos SNPs no haplogrupo E3b, que podem ocorrer, apesar de pouco frequentes (Publicação VII: Fernandes *et al.*, 2008a).

Os microsátélites do cromossoma Y permitiram definir a idade de coalescência de mutações, nomeadamente algumas deleções no gene *DAZ*, localizado na região AZFc do cromossoma Y, e concluir que muitas dessas deleções, supostamente associadas à infertilidade, são polimorfismos e terão sido o resultado de um processo evolutivo do cromossoma Y, paralelo à evolução observada nas mutações que definem os haplogrupos (Publicação X: Fernandes *et al.*, 2006).

Verificou-se a existência de Desequilíbrio de Ligação (*Linkage Disequilibrium*) entre dois microsátélites (D5S818 e CSF1PO) e dois polimorfismos do gene *IL4* (IL4RP2 e IL4590) localizados no cromossoma 5, tendo-se utilizado esses microsátélites para analisar a variabilidade dos dois polimorfismos e inferir sobre a sua ancestralidade e evolução permitindo assim justificar a diferença na frequência dos polimorfismos nas várias populações estudadas.

Foi possível esclarecer algumas dúvidas, através de estudos de LD, acerca da origem da elevada frequência do alelo T resultante da mutação pontual 590 C/T do gene *IL4*, bem como de outro polimorfismo no mesmo gene que consiste numa deleção de 70 pares de bases (IL4RP2) e que se encontra em LD com o primeiro, indicando a ocorrência de um crescimento rápido da população associado a uma selecção positiva que justifica a elevada frequência desta mutação em populações africanas (Publicação XI: Berenguer *et al.*, *submitted*).

O *locus* mais apropriado para ajudar a perceber a selecção que ocorreu neste caso, será o CSF1PO pois tem um maior valor de *Rst* e os *loci* que apresentem maiores valores de *Rst* são os candidatos mais prováveis a selecção. Quando esta ocorre em determinadas regiões contendo esses *loci* então outros *loci* na mesma região deverão apresentar também altos valores de *Rst* (Kayser *et al.*, 2003).

Apesar de elevados níveis de variabilidade em populações africanas indicarem que estas populações são mais antigas ou que têm um tamanho maior (Kayser *et al.*, 2003), o que se verifica para os alelos 183 e T do gene *IL4*, que apresentam uma maior variabilidade e aparecem com maior frequência nas populações africanas, parece dever-se a um crescimento rápido destas populações associado a uma possível selecção destes alelos, que se encontram em *LD*. Estas mutações, que são mais recentes do que as ancestrais (253 e C) que aparecem noutras espécies, aparecem com frequências altas nestas populações africanas (Publicação XI: Berenguer *et al.*, *submitted*).

4.5 PERSPECTIVAS FUTURAS

Com base nos trabalhos já desenvolvidos, outros estudos semelhantes poderão ser feitos para novos genes associados a estas ou outras doenças.

Seria interessante continuar o estudo da população do Arquipélago da Madeira, aumentando a amostragem de maneira a possibilitar a análise das populações de cada um dos concelhos.

Uma vez que o efeito da selecção parece ser mais acentuado no cromossoma X do que nos autossomas, seria interessante estudar os microsatélites do cromossoma X e comparar com os estudos já realizados para outros *loci*.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Adinolfi, M., Pertl, B., & Sherlock, J. (1997). Rapid detection of aneuploidies by microsatellite and the quantitative fluorescent polymerase chain reaction. *Prenatal Diagnosis*, 17(13), 1299-1311.
- Aguirell, E., Li, R., Rannug, U., Norming, U., Tribukait, B., & Ramel, C. (1992). Detection of DNA alterations in human bladder tumors by DNA fingerprint analyses. *Cancer Genetics and Cytogenetics* , 61(1), 53-60.
- Aiman, J., Griffin, J. E., Gazak, J. M., Wilson, J. D., & MacDonald, P. C. (1979). Androgen insensitivity as a cause of infertility in otherwise normal men. *New England Journal of Medicine* 300(5), 223-227.
- Alford, R. L., Hammond, H. A., Coto, I., & Caskey, C. T. (1994). Rapid and efficient resolution of parentage by amplification of Short Tandem Repeats. *American Journal of Human Genetics*, 55, 190-195.
- Amos, W., Sawcer, S. J., Feakes, R. W., & Rubinsztein, D. C. (1996). Microsatellites show mutational bias and heterozygote instability. *Nature Genetics*, 13, 390-391.
- Amos, W. (1999). A comparative approach to the study of microsatellite evolution. In D. B. Goldstein & C. Schlötterer (Eds.), *Microsatellites evolution and applications*. New York: Oxford University Press.
- Anderson, E. C., & Slatkin, M. (2004). Population-genetic basis of haplotype blocks in the 5q31 region. *American Journal of Human Genetics*, 74, 40-49.
- Armour, J. A. L., Alegre, S. A., Miles, S., Williams, L., & Badge, R. M. (1999). Minisatellites and mutation processes in tandemly repetitive DNA. In D. B. Goldstein & C. Schlötterer (Eds.), *Microsatellites evolution and applications* (pp. 24-32). New York: Oxford University Press.

- Armour, J. A. (2006). Tandemly repeated DNA: why should anyone care? *Mutation Research*, 598, 6-14.
- Ashley, C. T., & Warren, S. T. (1995). Trinucleotide repeat expansion and human disease. *Annual Reviews Genet*, 29, 703-728.
- Avery, P. J. (1984). The population genetics of haplo-diploids and X-linked genes. *Genetic Research* 44, 321-341.
- Bamshad, M. (2005). Genetic influences on health: does race matter? *Journal of the American Medical Association*, 294(8), 937-946.
- Bandelt, H.-J., Forster, P., Sykes, B. C., & Richards, M. B. (1995). Mitochondrial portraits of human populations. *Genetics*, 141, 743-753.
- Bandelt, H.-J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37-48.
- Bar, W. (1992). Recommendations of the DNA comission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms. *Forensic Science International*, 55, 1-3.
- Barcellos, C. J. (1899). *Subsídios para a história de Cabo Verde e Guiné*. Lisboa: Academia Real das Ciências.
- Beckman, J., & Weber, J. (1992). Survey of human and rat microsatellites. *Genomics*, 4(12), 627-631.
- Beleza, S., Alves, C., González-Neira, A., Lareu, M. V., Amorim, A., Carracedo, A., et al. (2003). Extending STR markers in Y chromosome haplotypes. *International Journal of Legal Medicine*, 117(1), 27-33.
- Beleza, S., Gusmão, L., Lopes, A., Alves, C., Gomes, I., Giouzeli, M., et al. (2006). Micro-phylogeographic and demographic history of Portuguese male lineages. *Annals of Human Genetics* 70(2), 181-194.
- Berenguer, A., Câmara, R., Brehm, A., & Fernandes, A. T. (2009). Distribution of polymorphisms IL4-590 C/T and IL4 RP2 in human populations from Madeira, Azores, Cape Verde and Guinea-Bissau. *American Journal of Human Biology*, (submitted).

- Bodmer, W. F. (1986). HLA today. *Human Immunology* 17(4), 490-503.
- Boehnke, M. (1994). Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes. *American Journal of Human Genetics* 55, 379-390.
- Bové, J. V. M. G., Cleton-Jansen, A. M., Kuipers-Dijkshoorn, N. J., Van Den Broek, L. J. C. M., Taminiau, A. H. M., Cornelisse, C. J., et al. (1999). Loss of heterozygosity and DNA ploidy point to a diverging genetic mechanism in the origin of peripheral and central chondrosarcoma. *Genes, Chromosomes and Cancer*(26), 237-246.
- Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R., & Cavalli-Sforza, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, 368, 455-457.
- Brehm, A., Pereira, L., Bandelt, H. J., Prata, M. J., & Amorim, A. (2002). Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Annals of Human Genetics*, 66, 49-60.
- Brehm, A., Pereira, L., Kivisild, T., & Amorim, A. (2003). Mitochondrial portraits of the Madeira and Açores archipelagos witness different genetic pools of its settlers. *Human Genetics* 114, 77-86.
- Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J., & Rolf, B. (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American Journal of Human Genetics*, 62, 1408-1415.
- Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., et al. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a proteina kinase family member. *Cell*, 4(68), 799-808.
- Budowle, B., Chakraborty, R., Giusti, A. M., Eisenberg, A. J., & Allen, R. C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *American Journal of Human Genetics*, 48(1), 137-144.

- Burger, J., Kirchner, M., Bramanti, B., Haak, W., & Thomas, M. G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences of the USA*, 104(10), 3736-3741
- Caglià, A., Novelletto, A., Dobosz, M., Malaspina, P., Ciminelli, B. M., & Pascali, V. L. (1997). Y-chromosome STR loci in Sardinia and continental Italy reveal islander-specific haplotypes. *European Journal of Human Genetics* 5(5), 288-292.
- Carita, R. (1989). *História da Madeira (1420-1566): povoamento e produção açucareira* (Vol. 1). Funchal: Governo Regional da Madeira.
- Carrington, M., Marti, D., Wade, J., Klitz, W., Barcellos, L., Thomson, G., et al. (1999). Microsatellite markers in complex disease: mapping disease-associated regions within the human major histocompatibility complex. In D. B. Goldstein & C. Schlotterer (Eds.), *Microsatellites evolution and applications*. (pp. 225-236). New York,: Oxford University Press.
- Caskey, C. T., Pizzuti, A., Fu, Y. H., Fenwick, R. G., & Nelson, D. L. (1992). Triplet repeat mutations in human disease. *Science*, 256, 784-789.
- Chakraborty, R., & Jin, L. (1992). Heterozygote deficiency, population substructure and their implications in DNA fingerprinting. *Human Genetics*, 88(3), 267-272.
- Chakraborty, R., Kimmel, M., Stivers, D. N., Davison, L. J., & Dekas, R. (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences of the USA*, 94, 1041-1046.
- Chamberlain, N. L., Driver, E. D., & Miesfeld, R. L. (1994). The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Research*, 22(15), 3181-3186.
- Chang, B. L., Zheng, S. L., Hawkins, G. A., Isaacs, S. D., Wiley, K. E., Turner, A., et al. (2002). Polymorphic GGC repeats in the androgen receptor gene are associated with hereditary and sporadic prostate cancer risk. *Human Genetics*, 110(2), 122-129.
- Chen, C., Lamharzi, N., Weiss, N. S., Etzioni, R., Dightman, D. A., Barnett, M., et al. (2002). Androgen receptor polymorphisms and the incidence of prostate cancer. *Cancer Epidemiology, Biomarkers & Prevention* 11(10), 1033-1040.

- Choong, C. S., Kemppainen, J. A., Zhou, Z. X., & Wilson, E. M. (1996). Reduced androgen receptor gene expression with first exon CAG repeat expansion. *Molecular Endocrinology*, 10(12), 1527-1535.
- Clayton, T. M., Whitaker, J. P., & Maguire, C. N. (1995). Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. *Forensic Science International*, 1(76), 7-15.
- Comings, D. E., Muhleman, D., J.P, J., & MacMurray, J. P. (2002). Parent-daughter transmission of the androgen receptor gene as an explanation of the effect of father absence on age menarche. *Child Development*, 73(4), 1046-1051.
- Cooper, D. L., Lahue, R. S., & Modrich, P. (1993). Methyl-directed mismatch repair is bidirectional. *Journal of Biological Chemistry*, 258(16), 11823-11829.
- Cooper, G., Amos, W., Hoffman, D., & Rubinsztein, D.C. (1996). Network analysis of human microsatellite haplotypes. *Human Molecular Genetics*, 5(11), 1759-1766.
- Coote, T., & Bruford, M. W. (1996). Human microsatellites applicable for analysis of genetic variation in apes and Old World monkeys. *Journal of Heredity*, 87, 406-410.
- Cornuet, J. M., & Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, 144, 2001-2014.
- Cummings, M. R. (2000). *Human heredity: principles and issues* (5th ed.). Pacific Grove: Brooks/Cole.
- Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., & Lander, E. S. (2001). High-resolution haplotype structure in the human genome. *Nature Genetics*, 29, 229-232.
- Di Rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M., & Freimer, N. B. (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the USA*, 91, 3166-3170.

- Di Rienzo, A., Donnelly, P., Toomajian, C., Sisk, B., Hill, A., Petzl-Erler, M. L., et al. (1998). Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. *Genetics*, 148, 1269-1284.
- Dupanloup, I., & Bertorelle, G. (2001). Inferring Admixture Proportions from Molecular Data: Extension to Any Number of Parental Populations. *Molecular Biology and Evolution*, 18(4), 672-675.
- Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., et al. (1993). Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genetics*, 4(4), 387-392.
- Edwards, A. I., Civitello, A., Hammond, H. A., & Caskey, C. T. (1991a). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics*, 49, 746-756.
- Edwards, M. C., Clemens, P. R., Tristan, M., Pizzuti, A., & Gibbs, R. A. (1991b). Pentanucleotide repeat length polymorphism at the human CD4 locus. *Nucleic Acids Research*, 19(17), 4792.
- Eisen, J. A. (1999). Mechanistic basis for microsatellite instability. In D. B. Goldstein & C. Schlotterer (Eds.), *Microsatellites evolution and applications* (pp. 34-48). New York: Oxford University Press.
- Elhaji, Y. A., Gottlieb, B., Lumbroso, R., Beitel, L. K., Foulkes, W. D., Pinsky, L., et al. (2001). The polymorphic CAG repeat of the androgen receptor gene: a potential role in breast cancer in women over 40. *Breast Cancer Research and treatment*, 70, 109-116.
- Estoup, A., Garnery, L., Solignac, M., & Cornuet, J. M. (1995). Microsatellite variation in honey bee (*Apis Mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics*, 140, 679-695.
- Estoup, A., & Cornuet, J. M. (1999). Microsatellite evolution: inferences from population data. In D. B. Goldstein & C. Schlotterer (Eds.), *Microsatellites evolution and applications* (pp. 49-64). New York: Oxford University Press.

- Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, 11, 591–1604.
- Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biological*, 3, 87-112.
- Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479-491.
- Excoffier, L., Laval, L. G., & Schneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1, 47-50.
- Faradz, S. M. H., Pattilha, M. Z., Leigh, D. A., Jenkins, M., Leggo, J., Buckley, M. F., et al. (2000). Genetic diversity at the FMR1 locus in the Indonesian population. *Annals of Human Genetics*, 64, 329-339.
- Ferlin, A., Bartoloni, L., Rizzo, G., Roverato, A., Garolla, A., & Foresta, C. (2004). Androgen receptor gene CAG and GGC repeat lengths in idiopathic male infertility. *Molecular Human Reproduction*, 10(6), 417-421.
- Fernandes, F. (2004). *Madeirenses Errantes*. Lisboa: Oficina do Livro.
- Fernandes, A. T., Brehm, A., Gusmão, L., & Amorim, A. (2001). Y-chromosome STR haplotypes in the Madeira archipelago population. *Forensic Science International*, 122, 178-180.
- Fernandes, A. T., & Brehm, A. (2002). Population data of five STRs in three regions from Portugal. *Forensic Science International*, 129, 72-74.
- Fernandes, A. T., Brehm, A., Alves, C., Gusmão, L., & Amorim, A. (2002). Genetic profile of the Madeira Archipelago population using the new Powerplex 16 system kit. *Forensic Science International*, 125(2-3), 281-283.
- Fernandes, A. T., & Brehm, A. (2003). Y-chromosome STR haplotypes in the Açores archipelago (Portugal). *Forensic Science International*, 135, 239-242.

- Fernandes, A. T., Velosa, R., Jesus, J., Carracedo, A., & Brehm, A. (2003). Genetic differentiation of the Cabo Verde archipelago population analysed by STR polymorphisms. *Annals of Human Genetics*, 67, 340-347.
- Fernandes, A. T., Gonçalves, R., Rosa, A., & Brehm, A. (2005). Analysis of Y chromosome and mtDNA variability in the Madeira Archipelago population. *International Congress Series, 21st Congress of International Society for Forensic Genetics, Açores*.
- Fernandes, A. T., Fernandes, S., Gonçalves, R., Sá, R., Costa, P., Rosa, A., et al. (2006). DAZ gene copies: evidence of Y chromosome evolution. *Molecular Human Reproduction*, 12(8), 519-523.
- Fernandes, A. T., Rosa, A., Gonçalves, R., Jesus, J., & Brehm, A. (2008a). The Y-chromosome short tandem repeats variation within haplogroup E3b: evidence of recurrent mutation in SNP. *American Journal of Human Biology*, 20(2), 185-190.
- Fernandes, A. T., Gonçalves, R., & Brehm, A. (2008b). Y-chromosome haplotype mismatch in different haplogroups: coincidence or evidence of SNP mutation? *Forensic Science International: Genetics, suppl 1*, 200-202.
- Fishel, R., & Wilson, T. (1997). MutS homologs in mammalian cells. *Current Opinion in Genetics & Development*, 7, 105-113.
- Fitzsimmons, N. N., Moritz, C., & Moore, S. S. (1995). Conservation and dynamics of microsatellite loci over 300 million years. *Molecular and Biology evolution*, 12(3), 432-440.
- Freedland, S. J., Aronson, W. J., Kane, C. J., Terris, M. K., Presti, J. C. J., Trock, B., et al. (2004). Biochemical outcome after radical prostatectomy among men with normal preoperative serum prostate-specific antigen levels. *Cancer*, 101(4), 748-753.
- Freitas, T., Brehm, A., & Fernandes, A. T. (2006). Frequency of the CCR5-Δ32 mutation in the atlantic island populations of Madeira, the Azores, Cabo Verde, and São Tomé e Príncipe. *Human Biology*, 78(6), 697-703.

- Fresco, J. R., & Alberts, B. M. (1960). The accommodation of noncomplementary bases in helical polyribonucleotides and deoxyribonucleic acids. *Proceedings of the National Academy of Sciences of the USA*, 46, 311-321.
- Frutuoso, G. (1977). *Saudades da Terra* (Vol. VI). Ponta Delgada.
- Fu, Y. H. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*, 67(6), 1047-1058.
- Gacy, A. M., Goellner, G., Juranić, N., Macura, S., & McMurray, C. T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell*, 81(4), 533-540.
- Gardner, H. (1993). *Multiple Intelligences: the theory in practice*. New York: Basic Books.
- Garza, J. C., & Freimer, N. B. (1996). Homoplasmy for size at microsatellite loci in humans and chimpanzees. *Genome Research*, 6, 211-217.
- Gill, P., Brenner, C., Brinkmann, B., Budowle, B., Carracedo, A., Jobling, M. A., et al. (2001). DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Forensic Science International*, 124, 5-10.
- Giovannucci, E., Stampfer, M. J., Krithivas, K., Brown, M., Dahl, D., Brufsky, A., et al. (1997). The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proceedings of the National Academy of Sciences of the USA*, 94(7), 3320-3323.
- Goldstein, D. B., & Clark, A. G. (1995). Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Research*, 23(19), 3882-3886.
- Goldstein, D. B., & Pollock, D. D. (1997). Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. *Journal of Heredity*, 88, 335-342.
- Goldstein, D. B. (2001). Islands of Linkage Disequilibrium. *Nature Genetics*, 29, 109-111.

- Gomes, I., Alves, C., Maxzud, K., Pereira, R., Prata, M., Sánchez-Diz, P., et al. (2007). Analysis of 10 X-STRs in three African populations. *Forensic Science International*, 1(2), 208-211.
- Gonçalves, R., Jesus, J., Fernandes, A. T., & Brehm, A. (2002). Genetic profile of a multi-ethnic population from Guiné-Bissau (west African coast) using the new Powerplex 16 system kit. *Forensic Science International*, 129, 78-80.
- Gonçalves, R., Rosa, A., Freitas, A. I., Fernandes, A. T., Kivisild, T., Villems, R., et al. (2003). Y-chromosome lineages in Cabo Verde islands witness the diverse geographic origin of its first male settlers. *Human Genetics*, 113, 467-472.
- Gonçalves, R., Freitas, A. I., Branco, M., Rosa, A., Fernandes, A. T., Zhivotovsky, L. A., et al. (2005). Y-chromosome lineages from Portugal, Madeira and Açores record elements of sephardim and berber ancestry. *Annals of Human Genetics*, 69, 443-454.
- Gourdon, G., Dessen, P., Lia, A. S., Junien, C., & Hofmann-Radvanyi, H. (1997). Intriguing association between disease associated unstable trinucleotide repeat and CpG island. *Annals of Genetics*, 2(40), 73-77.
- Grimaldi, M. C., & Crouau-Roy, B. (1997). Microsatellite allelic homoplasy due to variable flanking sequences. *Journal of Molecular Evolution*, 44, 336-340.
- Guill, J. H. (1993). *A history of the Azores Islands* Tulare CA: Division of Golden Shield International Publications Cooperation.
- Gunn, P. R., Trueman, K., Stapleton, P., & Klarkowski, D. B. (1997). DNA analysis in disputed parentage: the occurrence of two apparently false exclusions of paternity, both at short tandem repeat (STR) loci, in the one child. *Electrophoresis*, 18(9), 1650-1652.
- Gusmão, L., González-Neira, A., Sánchez-Diz, P., Lareu, M., Amorim, A., & Carracedo, A. (2000). Alternative primers for DYS391 typing: advantages of their application to forensic genetics. *Forensic Science International*, 112(1), 49-57.

- Gusmão, L., González-Neira, A., Alves, C., Lareu, M. V., Costa, S., Amorim, A., et al. (2002). Chimpanzee homologous of human Y specific STRs. A comparative study and a proposal for nomenclature. *Forensic Science International*, 126(136), 129-136.
- Gusmão, L., Sanchez-Diz, P., Calafell, F., Martin, P., Alonso, C. A., Álvarez-Fernandez, F., et al. (2005). Mutation rates at Y chromosome specific microsatellites. *Human Mutation*, 26(6), 520-528.
- Gusmão, L., Butler, J. M., Carracedo, A., Gill, P., Kayser, M., Mayr, W. R., et al. (2006). DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Science International*, 157, 187-197.
- Gusmão, L., Sánchez-Diz, P., Alves, C., Gomes, I., Zarrabeitia, M., Abovich, M., et al. (2009). A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. *International Journal of Legal Medicine*, 123(3), 227-234.
- Hakimi, J. M., Schoenberg, M. P., Rondinelli, R. H., Piantadosi, S., & Barrack, E. R. (1997). Androgen receptor variants with short glutamine or glycine repeats may identify unique subpopulations of men with prostate cancer. *Clinical Cancer Research*, 3(9), 1599-1608.
- Halling, K. C., French, A. J., McDonnell, S. K., Burgart, L. J., Schaid, D. J., Peterson, B. J., et al. (1999). Microsatellite instability and 8p allelic imbalance in stage B2 and C colorectal cancers. *Journal of the National Cancer Institute*, 91(15), 1295-1303.
- Hamada, H., Seidman, M., Howard, B. H., & Gorman, C. (1984). Enhanced gene expression by the poly(dT-dG) poly(dC-dA) sequence. *Molecular and Cellular Biology*, 4(12), 2622-2630.
- Hammer, M. F., Spurdle, A. B., Karafet, T., Bonner, M. R., Wood, E. T., Novelletto, A., et al. (1997). The geographic distribution of human Y chromosome variation. *Genetics*, 145, 787-805.

- Hammond, H. A., Jin, L., Zhong, Y., Caskey, C. T., & Chakraborty, R. (1994). Evaluation of 13 short tandem repeat loci for use in personal identification applications. *American Journal of Human Genetics*, 55, 175-189.
- Hancock, J. M. (1999). Microsatellites and other simple sequences: genomic context and mutational mechanisms In D. Goldstein & C. Schlötterer (Eds.), *Microsatellites: Evolution and Applications* (pp. 1-9). New York: Oxford University Press.
- Hardy, H. G. (1908). Mendelian proportions in a mixed population. *Science*, 28(706), 49-50.
- Harley, H. G., Rundle, S. A., MacMillan, J. C., Myring, J., Brook, J. D., Crow, S., et al. (1993). Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in Myotonic Dystrophy. *American Journal of Human Genetics*, 52, 1164-1174.
- Havill, L. M., & Dyer, T. D. (2005). Association mapping: methodologies, strategies, and issues. *Genetic Epidemiology*, 29(Suppl 1), S77-S85.
- Hiort, O., Holterhus, P. M., Hörter, T., Schulze, W., Kremke, B., Bals-Pratsch, M., et al. (2000). Significance of mutations in the androgen receptor gene in males with idiopathic infertility. *Journal of Clinical Endocrinology & Metabolism*, 85(8), 2810-2815.
- Hoff-Olsen, P., Meling, G. I., & Olaisen, B. (1998). Variation in mutation rate and direction between tetranucleotide STR loci in human colorectal carcinomas. *Annals of Human Genetics*, 62, 1-7.
- Hoff-Olsen, P., Mevåg, B., Staalstrøm, E., Hovde, B., Egeland, T., & Olaisen, B. (1999). Extraction of DNA from decomposed human tissue. An evaluation of five extraction methods for short tandem repeat typing. *Forensic Science International* 105(3), 171-183.
- Honchel, R., Halling, K. C., & Thibodeau, S. N. (1995). Genomic instability in neoplasia. *Seminars in Cell Biology*, 6(1), 45-52.
- Hsing, A., Tsao, L., & Devesa, S. (2000). International trends and patterns of prostate cancer incidence and mortality. *International Journal of Cancer*, 85(1), 60-67.

- Imbert, G., Kretz, C., Johnson, K. L., & Mandel, J. L. (1993). Origin of the expansion mutation in myotonic dystrophy. *Nature Genetics*, 4(1), 72-76.
- Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Hypervariable "minisatellite" regions in human DNA. *Nature*, 314, 67-73.
- Jeffreys, A. J., Allen, M. J., Hagelberg, E., & Sonnberg, A. (1992). Identification of the skeletal remains of Josef Mengele by DNA analysis. *Forensic Science International*, 56(1), 65-76.
- Jeffreys, A. J., Tamaki, K., MacLeod, A., Monckton, D. G., Neil, D. L., & Armour, J. A. (1994). Complex gene conversion events in germline mutation at human minisatellites. *Nature Genetics*, 6(2), 136-145.
- Jeffreys, A. J., & Neumann, R. (2002). Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nature Genetics*, 31, 267-271.
- Jobling, M., & Tyler-Smith, C. (1995). Fathers and sons: the Y chromosome and human evolution. *Trends in Genetics*, 11(11), 449-456.
- Jobling, M. A., & Tyler-Smith, C. (2000). New uses for new haplotypes, the human Y chromosome, disease and selection. *Trends in Genetics*, 16(8), 356-362.
- Jobling, M. A., Hurles, M. E., & Tyler-Smith, C. (2004). *Human Evolutionary Genetics: origins, people and disease* (1^a ed.). New York: Garland Publishing.
- Junge, A., Brinkmann, B., Fimmers, R., & Madea, B. (2006). Mutations or exclusion: an unusual case in paternity testing. *International Journal of Legal Medicine*, 120(6), 360-363.
- Kashi, Y., King, D., & Soller, M. (1997). Simple repeats as a source of quantitative genetic variation. *Trends in Genetics*, 13(2), 74-78.
- Kashi, Y., & Soller, M. (1999). Functional roles of microsatellites and minisatellites. In D. B. Goldstein & C. Schlötterer (Eds.), *Microsatellites evolution and applications*. New York: Oxford University Press.
- Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., et al. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nature Genetics*, 8(3), 221-228.

- Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., et al. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. *International Journal of Legal Medicine*, 110(3), 125-133.
- Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., et al. (2000). Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *American Journal of Human Genetics*, 66, 1580-1588.
- Kayser, M., Brauer, S., & Stoneking, M. (2003). A Genome scan to detect candidate regions influenced by local natural selection in human populations. *Molecular and Biology evolution*, 20(6), 893-900.
- Kimmel, M., Chakraborty, R., King, J. P., Bamshad, M., Watkins, W. S., & Jorde, L. B. (1998). Signatures of population expansion in microsatellite repeat data. *Genetics*, 148, 1921-1930.
- Kimura, M., & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, 49, 725-738.
- Kimura, M., & Ohta, T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the USA*, 75(6), 2868-2872.
- King, T. E., Parkin, E. J., Swinfield, G., Cruciani, F., Scozzari, R., Rosa, A., et al. (2007). Africans in Yorkshire? The deepest-rooting clade of the Y phylogeny within an English genealogy. *European Journal of Human Genetics*, 15, 288-293.
- Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I., & Khorana, H. G. (1971). Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. *Journal of Molecular Biology*, 56(2), 341-361.
- Kolodner, R. (1996). Biochemistry and genetics of eukaryotic mismatch repair. *Genes & Development*, 10, 1433-1442.
- Krawczak, M., V. Ball, E. V., & Cooper, D. N. (1998). Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. *American Journal of Human Genetics*, 63, 474-488.

- Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., et al. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science*, 252(5013), 1711-1714.
- Kunkel, T. (1992). DNA replication fidelity. *The Journal of Biological Chemistry*, 267(26), 18251-18254.
- Kunkel, T. (1993). Slippery DNA and disease. *Nature*, 365, 207-208.
- Kunst, C. B., Zerylnick, C., Karickhoff, L., Eichler, E., Bullard, J., Chalifoux, M., et al. (1996). FMR1 in global populations. *American Journal of Human Genetics*, 58(3), 513-522.
- Kunzler, P., Matsuo, K., & Schaffner, W. (1995). Pathological, physiological, and evolutionary aspects of short unstable DNA repeats in the human genome. *Biological Chemistry Hoppe Seyler*, 376(4), 201-211.
- La Spada, A., Wilson, E., Lubahn, D., Harding, A., & Fischbeck, K. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, 351(6330), 77-79.
- Lareu, M. V., Phillips, C. P., Carracedo, A., Lincoln, P. J., Syndercombe, C. D., & Thomson, J. A. (1994). Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. *Forensic Science International*, 66(1), 41-52.
- Lee, C., Wevrick, R., Fisher, R. B., Ferguson-Smith, M. A., & Lin, C. C. (1997). Human centromeric DNAs. *Human Genetics*, 100, 291-304.
- Levinson, G., & Gutman, G. A. (1987). Slipped-Strand Mispairing: a major mechanism for DNA sequence evolution. *Molecular and Biology Evolution*, 4(3), 203-221.
- Li, N., & Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, 165, 2213-2233.
- Liu, B., Farrington, S. M., Petersen, G. M., Hamilton, S. R., Parsons, R., Papadopoulos, N., et al. (1995). Genetic instability occurs in the majority of young patients with colorectal cancer. *Nature Medicine*, 1(4), 348-352.

- Lubahn, D. B., Joseph, D. R., Sullivan, P. M., Willard, H. F., French, F. S., & Wilson, E. M. (1988). Cloning of human androgen receptor complementary DNA and localization to the X chromosome. *Science*, 240(4850), 327-330.
- Luty, J. A., Guo, Z., Willard, H. F., Ledbetter, D. H., Ledbetter, S., & Litt, M. (1990). Five polymorphic microsatellite VNTRs on the human X chromosome. *American Journal of Human Genetics*, 46, 776-783.
- MacMillan, J. C., Snell, R. G., Tyler, A., Houlihan, G. D., Fenton, I., Cheadle, J. P., et al. (1993). Molecular analysis and clinical correlations of the Huntington's disease mutation. *Lancet*, 342(8877), 954-958.
- Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., & Haworth, I. A. (1995). Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)₁₅. *Nucleic Acids Research*, 23(6), 1050-1059.
- Morin, P. A., Wallis, J., Moore, J. J., & Woodruff, D. S. (1994). Paternity exclusion in a community of wild chimpanzees using hypervariable simple sequence repeats. *Molecular Ecology*, 3(5), 469-477.
- Moxon, E. R., Rainey, P. B., Nowak, M. A., & Lenski, R. E. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. *Current Biology*, 4(1), 24-33.
- Müller, I., Beeger, C., Alix-Panabières, C., Rebillard, X., Pantel, K., & Schwarzenbach, H. (2008). Identification of Loss of Heterozygosity on circulating free DNA in peripheral blood of prostate cancer patients: potential and technical improvements. *Clinical Chemistry*, 54, 688-696.
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposium on Quantitative Biology*, 51, 263-273.
- Nachman, M. W., & Crowell, S. L. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics*, 156, 297-304.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA*, 70(12), 3321-3323.
- Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.

- Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., et al. (1991). Instability of a 550 base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*, 252(5010), 1097-1102.
- Ohashi, J., & Tokunaga, K. (2003). Power of genome-wide linkage disequilibrium testing by using microsatellite markers. *Journal of Human Genetics*, 48, 487-491.
- Olaisen, B., Bar, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., et al. (1998). DNA recommendations 1997 of the International Society for Forensic Genetics. *Vox Sang*, 74(1), 61-63.
- Pakkala, S. (1991). DNA fingerprinting in the detection of residual disease in acute leukemia. *Leukemia*, 5(5), 437-440.
- Pardue, M., Lowenhaupt, K., Rich, A., & Nordheim, A. (1987). (dC-dA)_n.(dG-dT)_n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. *EMBO Journal*, 6(6), 1781-1789.
- Parker, B. O., & Marinus, M. G. (1992). Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the USA*, 89, 1730-1734.
- Payseur, B. A., Cutter, A. D., & Nachman, M. W. (2002). Searching for evidence of positive selection in the human genome using patterns of microatellite variability. *Molecular and Biology evolution*, 19(7), 1143-1153.
- Pemberton, J. M., Coltman, D. W., Coulson, T. N., & Slate, J. (1999). Using microsatellites to measure the fitness consequences of inbreeding and outbreeding. In D. B. Goldstein & C. Schlötterer (Eds.), *Microsatellites evolution and applications*. New York: Oxford University Press.
- Pereira, E. C. N. (1989). *Ilhas de Zarco* (4ª ed. Vol. 2). Funchal: Câmara Municipal do Funchal.
- Pereira, L., Prata, M. J., & Amorim, A. (2002). Mismatch distribution analysis of Y-STR haplotypes as a tool for the evaluation of identity-by-state proportions and

- significance of matches-European picture. *Forensic Science International*, 130, 147-155.
- Pertl, B., Weitgasser, U., Kopp, S., Kroisel, P. M., Sherlock, J., & Adinolfi, M. (1996). Rapid detection of trisomies 21 and 18 and sexing by quantitative fluorescent multiplex PCR. *Human Genetics*, 98(1), 55-59.
- Polymeropoulos, M. H., Xiao, H., Rath, D. S., & Merrill, C. R. (1991). Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). *Nucleic Acids Research*, 19(13), 3753.
- Primmer, C. R., Saino, N., Moller, A. P., & Ellegren, H. (1996). Directional evolution in germline microsatellite mutations. *Nature Genetics*, 13(4), 391-393.
- Prinz, M., Boll, K., Baum, H., & Shaler, B. (1997). Multiplexing of Y chromosome specific STRs and performance for mixed samples. *Forensic Science International*, 85(3), 209-218.
- Puers, C., Hammond, H. A., Jin, L., Caskey, C. T., & Schumm, J. W. (1993). Identification of repeat sequences heterogeneity at the polymorphic short tandem repeat locus HUMTH01 [AATG]_n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. *American Journal of Human Genetics*, 53, 953-958.
- Queller, D. C., & Goodnight, K. F. (1989). Estimation of genetic relatedness using allozyme data. *Evolution*, 43, 258-275.
- Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for Biologists*. New York: Cambridge University Press.
- Rannala, B., & Reeve, J. P. (2001). High-resolution multipoint Linkage-Disequilibrium mapping in the context of a Human Genome sequence. *American Journal of Human Genetics*, 69, 159-178.
- Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86, 248-249.
- Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, 105, 767-779.

- Richards, R. I., & Sutherland, G. R. (1997). Dynamic mutation: possible mechanisms and significance in human disease. *Trends in Biochemical Sciences*, 22, 432-436.
- Rico, C., Rico, I., & Hewitt, G. (1996). 470 million years of conservation of microsatellite loci among fish species. *Proceedings of the Royal Society of Biological Sciences*, 263(1370), 549-557.
- Risch, N. (2000). Searching for genes in complex diseases: lessons from systemic lupus erythematosus. *The Journal of Clinical Investigation*, 105(11), 1503-1506.
- Rogers, F. M. (1979). *Atlantic Islanders of the Azores and Madeira*. North Quincy, MA: The Christopher Publishing House.
- Rose, O., & Falush, D. (1998). A threshold size for microsatellite expansion. *Molecular Biology and Evolution*, 15(5), 613-615.
- Rousset, F., & Raymond, M. (1995). Testing heterozygote excess and deficiency. *Genetics*, 140(4), 1413-1419.
- Rubinsztein, D. C., Leggo, J., & Amos, W. (1995a). Microsatellite evolve more rapidly in humans than in chimpanzees. *Genomics*, 30, 610-612.
- Rubinsztein, D. C., Amos, W., Leggo, J., Goodburn, S., Jain, S., Li, S., et al. (1995b). Microsatellite evolution - evidence for directionality and variation in rate between species. *Nature Genetics*, 10, 337-343.
- Rubinsztein, D. C., Amos, B., & Cooper, G. (1999). Microsatellite and trinucleotide-repeat evolution: evidence for mutational bias and different rates of evolution in different lineages. *Philosophical Transactions of the Royal Society*, 354, 1095-1099.
- Russell-Wood, A. J. (1998). *The Portuguese empire, 1415-1808. A world on the move*. Baltimore: Johns Hopkins University Press.
- Samura, O., Sohda, S., Johnson, K. L., Pertl, B., Ralston, S., Delli-Bovi, L. C., et al. (2001). Diagnosis of trisomy 21 in fetal nucleated erythrocytes from maternal blood by use of short tandem repeat sequences. *Clinical Chemistry*, 47(9), 1622-1626.
- Santos, M. L., Sibov, T. T., Nishimoto, I. N., Kowaski, L. P., Miracca, E. C., & Nagai, M. A. (2004). The CAG repeat polymorphism in the androgen receptor gene (AR) and its relationship to head and neck cancer. *Oral Oncology*, 40, 177-182.

- Schlesselman, J. J. (1982). *Case-control studies: design, conduct and analysis*. Oxford: Oxford University Press.
- Schlötterer, C., & Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, 20(2), 211-215.
- Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109, 365-371.
- Schlötterer, C. (2002). Towards a molecular characterization of adaptation in local populations. *Current Opinion in Genetics & Development*, 12, 683-687.
- Schneider, P. M., Meuser, S., Waiyawuth, W., Seo, Y., & Rittner, C. (1998). Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. *Forensic Science International*, 97(1), 61-70.
- Schneider-Stock, R., Gunther, T., Roessner, A., & Epplen, J. T. (1998). Somatic DNA alterations in breast carcinomas of different lymph-node status by DNA fingerprint analyses. *Cancer Genetics and Cytogenetics*, 103, 149-154.
- Schwarzenbach, H., Müller, V., Beeger, C., Gottberg, M., Stahmann, N., & Pantel, K. (2007). A critical evaluation of loss of heterozygosity detected in tumor tissues, blood serum and bone marrow plasma from patients with breast cancer. *Breast Cancer Research*, 9, R66.
- Serra, E., Pros, E., garcia, C., López, E., Gili, M. L., Gómez, C., et al. (2007). Tumor LOH analysis provides reliable linkage information for prenatal genetic testing of sporadic NF1 patients. *Genes Chromosomes Cancer*, 46(9), 820-827.
- Serrão, J. V. (1980). *História de Portugal* (Vol. 2). Lisboa: Verbo.
- Shriver, M. D., Jin, L., Chakraborty, R., & Boerwinkle, E. (1993). VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics*, 134, 983-993.
- Sia, E. A., Kokoska, R. J., Dominska, M., Greenwell, P., & Petes, T. D. (1997). Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Molecular and Cellular Biology*, 17(5), 2851-2858.

- Silva, F. A., & Meneses, C. A. (1984). *Elucidário madeirense*. Funchal: Direcção Regional dos Assuntos Culturais.
- Slatkin, M. (1994). Linkage Disequilibrium in growing and stable populations. *Genetics*, 137, 331-336.
- Smith, G. P. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science*, 191(4227), 528-535.
- Snell, R. G., MacMillan, J. C., Cheadle, J. P., Fenton, I., Lazarou, L. P., Davies, P., et al. (1993). Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nature Genetics*, 4(4), 329-330.
- Spínola, H., Middleton, D., & Brehm, A. (2005). HLA genes in Portugal inferred from sequence-based typing: in the crossroad between Europe and Africa. *Tissue Antigens*, 66, 26-36.
- Stallings, R. L. (1994). Distribution of Trinucleotide Microsatellites in Different Categories of Mammalian Genomic Sequence: Implications for Human Genetic Diseases. *Genomics*, 21(1), 116-121.
- Stanford, J. L., Just, J. J., Gibbs, M., Wicklund, K. G., Neal, C. L., Blumenstein, B. A., et al. (1997). Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. *Cancer Research*, 57(6), 1194-1198.
- Stephens, J. C., Smith, M. W., Shin, H. D., & O'Brien, J. (1999). Tracking linkage disequilibrium in admixed populations with MALD using microsatellite loci. In D. B. Goldstein & C. Schlotterer (Eds.), *Microsatellites evolution and applications*. New York: Oxford University Press.
- Strand, M., Prolla, T. A., Liskay, R. M., & Petes, T. D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*, 365, 274-276.
- Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., & Terzaghi, E. (1966). Frameshift mutations and the genetic code. *Cold Spring Harbor Symposium on Quantitative Biology*, 31, 77-84.

- Stumpf, M. P., & Goldstein, D. B. (2001). Genealogical and evolutionary inference with the human Y chromosome. *Science*, 291(5509), 1738-1742.
- Stumpf, M. P., & Goldstein, D. B. (2003). Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium. *Current Biology*, 13(1), 1-8.
- Suter, N. M., Malone, K. E., Daling, J. R., Doody, D. R., & Ostrander, E. A. (2003). Androgen receptor (CAG)_n and (GGC)_n polymorphisms and breast cancer risk in a population-based case-control study of young women. *Cancer Epidemiology, Biomarkers and Prevention*, 12, 127-135.
- Sutherland, G. R., & Richards, R. I. (1995). Simple tandem DNA repeats and human genetic disease. *Proceedings of the National Academy of Sciences of USA*, 92, 3636-3641.
- Tachida, H., & Iizuka, M. (1992). Persistence of repeated sequences that evolve by replication slippage. *Genetics*, 131, 471-478.
- Tautz, D., & Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, 12(10), 4127-4138.
- Thibodeau, S. N., Bren, G., & Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science*, 260(5109), 816-819.
- Thibodeau, S. N., French, A. J., Roche, P. C., Cunningham, J. M., Tester, D. J., Lindor, N. M., et al. (1996). Altered expression of hMSH2 and hMLH1 in tumors with microsatellite instability and genetic alterations in mismatch repair genes. *Cancer Research*, 56, 4836-4840.
- Thibodeau, S. N., French, A. J., Cunningham, J. M., Tester, D., Burgart, L. J., Roche, P. C., et al. (1998). Microsatellite instability in colorectal cancer: different mutate phenotypes and the principal involvement of hMLH1. *Cancer Research*, 58, 1713-1718.
- Valdes, A. M., Slatkin, M., & Freimer, N. B. (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*, 133, 737-749.
- Vandenberg, N., van Oorschot, R. A., & Mitchell, R. J. (1997). An evaluation of selected DNA extraction strategies for short tandem repeat typing. *Electrophoresis*, 18(9), 1624-1626.

- Velosa, R., Fernandes, A. T., & Brehm, A. (2002). Genetic profile of the Açores archipelago population using the new PowerPlex 16 system kit. *Forensic Science International*, 129, 68-71.
- Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., et al. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, 65(5), 905-914.
- Vieira, A. (1991). *Os escravos no arquipélago da Madeira - séculos XV a XVII* (1ª ed.). Funchal: DRAC.
- Vigilant, L. (1999). An evaluation of techniques for the extraction and amplification of DNA from naturally shed hairs. *Journal of Biological Chemistry*, 380(11), 1329-1331.
- Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*, 10, 506-513.
- Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids. *Nature*, 4356, 737.
- Watterson, G. A. (1978). The Homozygosity Test of Neutrality. *Genetics*, 88, 405-417.
- Weber, J. L., & Wong, C. (1993). Mutation of human short tandem repeats. *Human Molecular Genetics*, 2(8), 1123-1128.
- Weir, B. S. (1996). *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sunderland, MA: Sinauer Associates.
- Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., et al. (1992). A second-generation linkage map of the human genome. *Nature*, 359(6398), 794-801.
- Werely, C. J., Heyns, C. F., Van Velden, D. J., & Van Helden, P. D. (1996). DNA fingerprint detection of somatic mutations in benign prostatic hyperplasia and prostatic adenocarcinoma. *Genes Chromosomes Cancer*, 17(1), 31-36.

- Westberg, L., Melke, J., Landen, M., Nilsson, S., Baghaei, F., Rosmond, R., et al. (2003). Association between a dinucleotide repeat polymorphism of the estrogen receptor alpha gene and personality traits in women. *Molecular Psychiatry*, 8(1), 118-122.
- Wierdl, M., Dominska, M., & Petes, T. D. (1997). Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics*, 146, 769-779.
- Willuweit, S., & Roewer, L. (2007). Y chromosome haplotype reference database (YHRD): Update. *Forensic Science International: Genetics*, 1(2), 83-87.
- Xu, X., Peng, M., Fang, Z., & Xu, X. (2000). The direction of microsatellite mutations is dependent upon allele length. *Nature Genetics*, 24, 396-399.
- Yepes, J. C. (2006). *Matemáticas aplicadas a la genética forense*. Madrid: Ministério del Interior.
- Zhivotovsky, L. A., Bennett, L., Bowcock, A., & Feldman, M. W. (2000). Human population expansion and microsatellite variation. *Molecular and Biology evolution*, 17(5), 757-767.
- Zhivotovsky, L. A., Underhill, P. A., Cinnioğlu, C., Kayser, M., Morar, B., Kivisild, T., et al. (2004). The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *American Journal of Human Genetics*, 74(1), 50-61.

6. PUBLICAÇÕES

- I. **Fernandes, A.T.**, Brehm, A., Alves, C., Gusmão, L. & Amorim, A. (2002) Genetic profile of the Madeira Archipelago population using the new PowerPlex 16 system kit. *Forensic Science International*, 125, 281-283.
- II. **Fernandes, A.T.**, Brehm, A., Gusmão, L. & Amorim, A. (2001) Y-chromosome STR haplotypes in the Madeira archipelago population. *Forensic Science International*, 122, 178-180.
- III. Velosa, R., **Fernandes, A.T.** & Brehm, A. (2002) Genetic profile of the Açores archipelago population using the new PowerPlex 16 system kit. *Forensic Science International*, 129, 68-71.
- IV. **Fernandes, A.T.** & Brehm, A. (2003) Y-chromosome STR haplotypes in the Açores archipelago (Portugal). *Forensic Science International*, 135, 239-242.
- V. **Fernandes, A.T.**, Velosa, R., Jesus, J., Carracedo, A. & Brehm, A. (2003) Genetic differentiation of the Cabo Verde archipelago population analyzed by STR polymorphisms. *Annals of Human Genetics*, 67, 340-347.
- VI. **Fernandes, A.T.** & Brehm, A. (2002) Population data of five STRs in three regions from Portugal. *Forensic Science International*, 129, 72-74.
- VII. Gonçalves, R., Jesus, J., **Fernandes, A.T.** & Brehm, A. (2002) Genetic profile of a multi-ethnic population from Guiné-Bissau (west African coast) using the new Powerplex 16 system kit. *Forensic Science International*, 129, 78-80.
- VIII. **Fernandes, A.T.**, Rosa, A., Gonçalves, R., Jesus, J. & Brehm, A. (2008a) The Y-chromosome short tandem repeats variation within haplogroups E3b: evidence of recurrent mutation in SNP. *American Journal of Human Biology*, 20, 185-190.

- IX. **Fernandes, A. T.**, Gonçalves, R. & Brehm, A (2008b) Y-chromosome haplotype mismatch in different haplogroups: coincidence or evidence of SNP mutation? *Forensic Science International Genetics*, Suppl1, 200-202.
- X. **Fernandes, A.T.**, Fernandes, S., Gonçalves, R., Sá, R., Costa, P., Rosa, A., Ferrás, C., Sousa, M., Brehm, A. & Barros, A. (2006) DAZ gene copies: evidence of Y chromosome evolution. *Molecular Human Reproduction*, 12, 519-523.
- XI. Berenguer, A., Câmara, R., Brehm, A. & **Fernandes, A. T.** (2009) Distribution of polymorphisms IL4 -590 C/T and IL4 – RP2 in the human populations of Madeira, Azores, Portugal, Cape Verde and Guinea-Bissau. *American Journal of Human Biology* (submitted).